

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2002, 13:11:02 ; Search time 4399.1 Seconds

(without alignments)
11520.387 Million cell updates/sec

Title: US-09-697-089-3

Perfect score: 3072
Sequence: 1 atgaatttcataagacaa.....ctttaactagtaactgct 3072

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

GenEmbl:*

1: gb_ba:*
2: gb_hlg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vi:*
30: em_higo_hum:*
31: em_higo_inv:*
32: em_higo_rod:*
33: em_hlg_hum:*
34: em_hlg_inv:*
35: em_hlg_rod:*
36: em_hlg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3072	100.0	3133	9	AY032589	AY032589 Homo sapi
2	3067.2	99.8	3219	9	AY035391	AY035391 Homo sapi
3	3065.6	99.8	3396	9	AY027787	AY027787 Homo sapi
4	3065.6	99.8	3581	9	AF376061	AF376061 Homo sapi
5	2001.6	65.2	160583	2	AC010968	AC010968 Homo sapi
6	1998.4	65.1	138909	9	CNS010D3	AL121653 BAC sequ
7	1316.4	42.9	13355	9	IR2005417	AL1289934 Homo sapi
8	815.4	26.5	1395	9	AY027788	AY027788 Homo sapi
9	461	15.0	768	9	AY027789	AY027789 Homo sapi
10	292.2	9.5	185281	2	AC011232	AC011232 Homo sapi
11	292	8.8	576	9	CNS010D8	AL121658 BAC sequ
12	269	8.8	578	9	AY027790	AY027790 Homo sapi
13	170	5.5	553	11	G55568	G55568 SHGC-100923
14	68.4	2.2	117791	10	AF242431S1	AF242431 Mus muscu
15	67.8	2.2	4752	10	AF135490	AF135490 Mus muscu
16	67.8	2.2	4815	10	AF102871	AF102871 Mus muscu
17	67.8	2.2	4823	10	AF135489	AF135489 Mus muscu
18	67.8	2.2	5362	10	AF135491	AF135491 Mus muscu
19	67.8	2.2	5497	10	AF007769	AF007769 Mus muscu
20	67.8	2.2	90650	10	AF242431S2	AF242432 Mus muscu
21	66.8	2.2	6829	10	AF135494	AF135494 Mus muscu
22	66.2	2.2	179252	10	AF131205	AF131205 Mus muscu
23	65.2	2.1	41613	10	AF242433S1	AF242433 Mus muscu
24	64.6	2.1	14784	10	AF242433S3	AF242435 Mus muscu
25	62	2.0	5269	10	AF135492	AF135492 Mus muscu
26	59	1.9	5366	6	E23944	E23944 Excessive o
27	59	1.9	5366	6	E24990	E24990 Apoptosis-1
28	59	1.9	5502	6	A64509	A64509 Sequence 1
29	59	1.9	5984	6	E23943	E23943 Excessive o
30	59	1.9	5984	6	E24989	E24989 Apoptosis-1
31	59	1.9	6124	6	A64529	A64529 Sequence 21
32	59	1.9	6124	9	HSU19251	U19251 Homo sapien
33	59	1.9	6133	6	A64510	A64510 Sequence 2
34	59	1.9	6228	3	A64531	A64531 Sequence 23
35	59	1.9	82976	9	AC005031	AC005031 Homo sapi
36	59	1.9	131078	9	HSU80017	U80017 Homo sapien
37	59	1.9	135306	2	AC010272	AC010272 Homo sapi
38	59	1.9	155809	2	AC012369	AC012369 Homo sapi
39	59	1.9	155974	9	AC022119	AC022119 Homo sapi
40	59	1.9	168814	9	AC010237	AC010237 Homo sapi
41	59	1.9	190871	9	AC044797	AC044797 Homo sapi
42	56.4	1.8	716	9	HSNA1P1	U21913 Human neuro
43	44.2	1.4	1141	6	AX083744	AX083744 Sequence
44	44	1.4	177639	2	AC091756	AC091756 Sus scrof
45	42.8	1.4	158082	2	AC091793	AC091793 Felis cat

ALIGNMENTS

RESULT	1	ALIGNMENTS
AY032589	AY032589	3133 bp mRNA
LOCUS	AY032589	Homo sapiens caspase recruitment domain protein 12 mRNA, complete cds.
DEFINITION	AY032589.1	GI:13899172
ACCESSION	AY032589	
VERSION	AY032589.1	GI:13899172
KEYWORDS		
SOURCE		human.
ORGANISM		Homo sapiens
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS		Geddes,B.J., Wang,L., Huang,W.-J., Lavellée,M., Manji,G.A., Brown,M., Jurman,M., Morganstern,J., Merriam,S., Gluckmann,A., Distefano,P.S. and Bertin,J.
TITLE		Human CARD12 is a Novel CED4/Apaf-1 Family Member That Induces Apoptosis
JOURNAL		Biochem. Biophys. Res. Commun. 284 (1), 77-82 (2001)
PUBMED		11374873

Oy 1501 gccaccagagctgttatgaagacctcgcagagtgatcaaacagcgttccttcgga 1560
 |||||
 Db 1536 gccaccagagcgttatgaagacactccgacagctgtatcaaacagcgttccttcgga 1595
 |||||
 Oy 1561 ctctccatcgcgaagagcctctctcgtgacaggaatctctgcaagtgtaaaaaaac 1620
 |||||
 Db 1596 cttttccatgcgcgaagagcctctctcgtgacaggaatctctgcaagtgtaaaaaaac 1655
 |||||
 Oy 1621 actgagcaagaatctcgaagccataacatcaatcccttctgtagtgatggatccat 1680
 |||||
 Db 1656 actgagcaagaatctcgaagccataacatcaatcccttctgtagtgatggatccat 1715
 |||||
 Oy 1681 ttatacaagagagtagatcccaatccagccctgagcccaagaattggaagcttcttcaa 1740
 |||||
 Db 1776 ttatatcaagagagtagatcccaatccagccctgagcccaagaattggaagcttcttcaa 1775
 |||||
 Oy 1741 ggtaaaagcttatatacaactcaagagacatccccgaattacttaattgacttcttgaa 1800
 |||||
 Db 1776 ggtaaaagcttatatacaactcaagagacatccccgaattacttaattgacttcttgaa 1835
 |||||
 Oy 1801 cattgcccgaattgtgcaagtgctctggaatcattaaactggaacttctatggggagct 1860
 |||||
 Db 1836 cattgcccgaattgtgcaagtgctctggaatcattaaactggaacttctatggggagct 1895
 |||||
 Oy 1861 atggtctcatgggaaagagctgcgaagaacacaggtggaatccacatgaaagaagcccca 1920
 |||||
 Db 1896 atggtctcatgggaaagagctgcgaagaacacaggtggaatccacatgaaagaagcccca 1955
 |||||
 Oy 1921 gaaacctcaatctcccaagagagctgtatccttcttctcaactggaagcaagaatctcag 1980
 |||||
 Db 1956 gaaacctcaatctcccaagagagctgtatccttcttctcaactggaagcaagaatctcag 2015
 |||||
 Oy 1981 actctggaagtgcaacatcccggaattcagcaagtggaatgaagaagacatacatatctg 2040
 |||||
 Db 2016 actctggaagtgcaacatcccggaattcagcaagtggaatgaagaagacatacatatctg 2075
 |||||
 Oy 2041 gggaaaatattcagctctgcgaagcctcagagctgcgaatataaagaatgtgtgtgtg 2100
 |||||
 Db 2076 gggaaaatattcagctctgcgaagcctcagagctgcgaatataaagaatgtgtgtgtg 2135
 |||||
 Oy 2101 gctggaagcctcagtgctgtcctcagcagcctgtgaagaacattatcctcctcagtgga 2160
 |||||
 Db 2136 gctggaagcctcagtgctgtcctcagcagcctgtgaagaacattatcctcctcagtgga 2195
 |||||
 Oy 2161 gccagctccctccacatagaagatgagagacacatccatctgttaacaaacctgaaacc 2220
 |||||
 Db 2196 gccagctccctccacatagaagatgagagacacacatccatctgtgttaacaaacctgaaacc 2255
 |||||
 Oy 2221 ttgaattatctgacctcagaatcaaacagctgcgggtgtgtcctcagcagagcttggt 2280
 |||||
 Db 2256 ttgaattatctgacctcagaatcaaacagctgcgggtgtgtcctcagcagagcttggt 2315
 |||||
 Oy 2281 aactggaagacactacaaagctcaatggaatacaataaagaatgaagaagatgct 2340
 |||||
 Db 2316 aactggaagacactacaaagctcaatggaatacaataaagaatgaagaagatgct 2375
 |||||
 Oy 2341 ataaactagctgaagccttgaaaaacctggaagaagatgtgttatttcaattgagccac 2400
 |||||
 Db 2376 ataaactagctgaagccttgaaaaacctggaagaagatgtgttatttcaattgagccac 2435
 |||||
 Oy 2401 ttgctgcatctggaagagagatgatatcaatgacagctcctgtcgaagtgaacccctgt 2460
 |||||
 Db 2436 ttgctgcatctggaagagagatgatatcaatgacagctcctgtcgaagtgaacccctgt 2495
 |||||
 Oy 2461 gacctggaagaatcaatagctcctgtcgtgtcgtcgaatgagtgaaatcccta 2520
 |||||
 Db 2496 gacctggaagaatcaatagctcctgtcgtgtcgtcgaatgagtgaaatcccta 2555
 |||||
 Oy 2521 gctcgaatcttcaaatctgtcgaactgagcatctctgatttatacgaataatcactg 2580
 |||||
 Db 2556 gctcgaatcttcaaatctgtcgaactgagcatctctgatttatacgaataatcactg 2615
 |||||

Oy 2581 gaaaaagaatggaatgaagcctctcatgactgatacgaagatgaacgtgtcagaacag 2640
 |||||
 Db 2616 gaaaaagaatggaatgaagcctctctcatgactgatacgaagatgaacgtgtcagaacag 2675
 |||||
 Oy 2641 ctcaacgcatagtatgctccctggggcgtgtgacgtgtcaagagcagcctgagagcctgtg 2700
 |||||
 Db 2676 ctcaacgcatagtatgctccctggggcgtgtgacgtgtcaagagcagcctgagagcctgtg 2735
 |||||
 Oy 2701 aaacattggggaggggtccccaactcgtcaagctgtggttgaaaaactgtgaactcaca 2760
 |||||
 Db 2736 aaacattggggaggggtccccaactcgtcaagctgtggttgaaaaactgtgaactcaca 2795
 |||||
 Oy 2761 gatacagagatgaatctttagtgatcttcttgaaagaacccctgaaaacctccag 2820
 |||||
 Db 2796 gatacagagatgaatctttagtgatcttcttgaaagaacccctgaaaacctccag 2855
 |||||
 Oy 2821 caagtgaatttggcgggaaatcgtgtgacagtgatgatatgcttgcctcatggtgtga 2880
 |||||
 Db 2856 caagtgaatttggcgggaaatcgtgtgacagtgatgatatgcttgcctcatggtgtga 2915
 |||||
 Oy 2881 ttggaatccttaagcaatgagtggttttctgacttagtactaagaatctcactgat 2940
 |||||
 Db 2916 ttggaatccttaagcaatgagtggttttctgacttagtactaagaatctcactgat 2975
 |||||
 Oy 2941 ccagcatatgacgaagaacttagccaagtgatcccaagtgatcacttcttcgaagaagct 3000
 |||||
 Db 2976 ccagcatatgacgaagaacttagccaagtgatcccaagtgatcacttcttcgaagaagct 3035
 |||||
 Oy 3001 agcctgtgtgtgcaatctgatatgatatcagtgatcagtgatcagtgatcagtgatcag 3060
 |||||
 Db 3036 agcctgtgtgtgcaatctgatatgatatcagtgatcagtgatcagtgatcagtgatcag 3095
 |||||
 Oy 3061 ctgagtaactgct 3072
 |||||
 Db 3096 ctgagtaactgct 3107
 |||||

RESULT 2
 AY035391
 LOCUS 3219 bp mRNA
 DEFINITION Homo sapiens ICB-protease activating factor mRNA, complete cds.
 ACCESSION AY035391
 VERSION AY035391.1 GI:14334214
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 3219)
 Poyet, J.L., Srinivasula, S.M., Tian, M., Razmara, M.,
 Fernandes-Alnemri, T. and Alnemri, E.S.
 Identification of Ipa1, a human caspase-1-activating protein
 related to Apaf-1
 J. Biol. Chem. 276 (30), 28309-28313 (2001)
 TITLE JOURNAL
 MEDLINE 21359454
 PUBMED 11390368
 REFERENCE 2 (bases 1 to 3219)
 Poyet, J.-L., Srinivasula, S.M., Fernandes-Alnemri, T. and
 Alnemri, E.S.
 Direct Submission
 Submitted (16-MAY-2001) Microbiology and Immunology, Thomas
 Jefferson University, 233 S. 10th Street, Philadelphia, PA 19107,
 USA

FEATURES
 source
 1..3219
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /map="p21-22"
 145..3219
 /note="IPAF; CED4/Apaf-1 family member; caspase-associated
 recruitment domain containing protein"
 /codon_start=-1

Db	1885	GGTAAAGCTTATATATATCAACCTAGGAGACATCCCGATTACTTATTTGACTTCTTTGAA	1944
Qy	1801	catttggccaatttgtgcaagtgctctcgtacattcatcaacctggaacttattatgvggagct	1860
Db	1945	CATTGGCCCAATTGTGCAAGTGGCCCTGGACCTCATTTAACTGGACTTTATGGGGACT	2004
Qy	1861	atggtctcatgvgaaagagctctcagaagacacagctgaaatccacatgvaagagcccca	1920
Db	2005	ATGCGCTTCATGGGGAAGAGCGCTGCAGAAAGACACAGGCGAATCCACATGGAAAGGCCCA	2064
Qy	1921	gaaacctaacctcccgcaagagctgtatcttgttcttaacctggaaggaagaaatcagg	1980
Db	2065	GAAACCTCACATCCACACAGGCGTGTATCTTGTTCCTCACTGGAAAGCAGGAATTTACGG	2124
Qy	1981	actctggaagctcaacaccccggaattccaagaagtgtaaatgaagaagatcaacatctg	2040
Db	2125	ACTCTGGAAGCTCACACCCCGGAAATTTCAGCAAGTTAAATACCAAGATATTCAGATATCTG	2184
Qy	2101	gctggaagcctcaagtttgttccctcagcaacctgtgaagaacattatctctcatgtgtgaa	2160
Db	2245	GCTGGAAAGCCTCAGTTTGGTGCCTCACACACTGTAAACAAATTATTTCTCTCATGGGTGAA	2304
Qy	2161	gcgaagctccctcaaccatagaagaatgagagacatatcaactcgtaaacaaacttgaaac	2220
Db	2305	GCCAGTCCCTCACCATAGAAAGATGAGAGGCAATCATCTGTATACCAACTGAAAC	2364
Qy	2221	ttgagtaattcatgacctacacgaatacaacgagctgcgggtggtctcagctgacagccttggt	2280
Db	2365	TTGAGTATTCATGACCTACAGATTAACAGGCTGCCGGTGGTCTGACTGACAGCTTTGGGT	2424
Qy	2281	aacttgaagaaccttacaaagctcataatgataaataagaatgaatgaagaagatgct	2340
Db	2425	AACCTGGAAGAACCTTACAAAGGCTCATTAATGATTAACATTAAGATGAATGAAGAAATGCT	2484
Qy	2341	ataaacttaagctgaagagccctcgaaaaaccctgaagaagaatggtttatattcaatttgccac	2400
Db	2485	ATAAATACTAGCTGAAGGCTCGAAAAACCTGAAAGAGATGTGTTTATTTTCATTGTGACCAAC	2544
Qy	2401	ttgtctgacacttggagagggaaatgataatcatagtaagtcctctgtlcaagtgaaacctgt	2460
Db	2545	TTGTCTGCATTTGGAGAGGGGAATGGAATTCATATGTCAGTCTGTGCAAGTGAACCTG	2604
Qy	2461	gaacctgaagaagaattcaatagtcctccctgctgtgtctgtcaaatgcaatgaaatccta	2520
Db	2605	GACCTTGAAGAAATTCAAATTATAGTCTCGTGTGCTTGTGCAAAATGCAAGTGAATTCCTA	2664
Qy	2521	gctagaagatctcaagaatttggttccaacatgagatctctgatttttcgaaaaattactgt	2580
Db	2665	GCTGAGATCTTCACAATTTTGGTCAAACTGAGGATTTCTTGATTTATCAGAAAAATTACCTG	2724
Qy	2581	gaaaaagaatgaaatgaaagctctctcaatgacatgatacagaagatgaacgtgctlaagacag	2640
Db	2725	GAAAAAGATGGAATGGAAGCTCTTCATGAACGTATGACAGAGATGAACCTGCTTAACACAG	2784
Qy	2641	ctcaacgcaactgaatgctgcctctggggctgtgacgtgtgcaagcagccttgaagcagcctgtg	2700
Db	2785	CTCACCCGACAGTATGCTGCTCGTGGGCTGTGAAGTCAAGCAGCCTGAGCAGCCTGTTGG	2844
Qy	2701	aaacatttgaagagaggtcccaacaactcgtccaagcttggtgtgaaaaaactgvgagactaca	2760
Db	2845	AAACATTTTGGAGGGAGGTCACCAACTGTCTACACTCTTGAGTTGAAAACCTGGAGACTCACA	2904
Qy	2761	gatacagagattgaagaattcaggatgcaatttttggaaagaaccccttgaaaaaacttccag	2820
Db	2905	GATACAGAGATTGAATTTTATAGTGCATTTTGTGGAAAGAACCCCTGTGAANAACCTTCCAG	2966
Qy	2821	caattgaatttgccgggaaatcgltgtagcagtgatgaaatgagcttgaccttcaatgggtgta	2880

[illegible]

KNMRLTDEIRILGAFEGKNPLKNFOQLNLAGNRSSDGLAFWJVENIKOLVFEDE
STKEPLDPAIVLRKLSOVLSTKTFLOEARLVGMQFDDDLSTVTFMKLVTA
BASE COUNT 992 a 737 c 793 g 874 t
ORIGIN

Query Match 99.8%; Score 3065.6; DB 9; Length 3336;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 3068; Conservative 0; Mismatches 4; Indels 3; Gaps 0;

OY 1 atgaattcataaaggacataagccgagcccttatccaagaagtgaagtactatata 60
DB ATGAATTTCTTAAGGACATATAGCCGAGCCCTTATTCAAAGAAATGGGAATGACTGTTATA 336
OY 61 aagcaaatccagatgaacctatgtatgaatgtctgaatcggaagaagtaacatc 120
DB AAGCAATATCAGATGACCTATTGTATGSAATGTTCTGAAATGCCGAAGAATNAICATC 336
OY 337 AAGCAATATCAGATGACCTATTGTATGSAATGTTCTGAAATGCCGAAGAATNAICATC 336
OY 121 attgctgcgaagaagtggagcagaatgctgctagaaggagatcaatgatattgaaa 180
DB ATTGCTGCAGAGAGGTGAGAGAGATGCTGTAAGAGGATCTTCATGATGATTTTGA 456
OY 397 ATTGCTGCAGAGAGGTGAGAGAGATGCTGTAAGAGGATCTTCATGATGATTTTGA 456
OY 181 aaggtctcagaagtcctgtlaacctcttctaatacccttaagaagtgaatctctcta 240
DB AAGGCTTCAGAGTCTGTATACCTCTTCTTAATCCCTTAAGAGATGGAATATACCTCTA 516
OY 457 AAGGCTTCAGAGTCTGTATACCTCTTCTTAATCCCTTAAGAGATGGAATATACCTCTA 516
OY 241 tctcaagacttgaatggacaagtcttctcatcagaatcagaagaagacttggacat 300
DB TTTTCAGACTTGAATGACAAAGCTTTTTCATCAGACATCAGAAAGAGACTTGGACGAT 517
OY 517 TTTTCAGACTTGAATGACAAAGCTTTTTCATCAGACATCAGAAAGAGACTTGGACGAT 517
OY 301 ttggctcagaatttaagaagactgtlaacacacacacacacacacacacacacacac 360
DB TTGGCTCAGACTTTAAAGGACTTGTACCAATCCCATCTTTTCTGAACCTTTTATTC 636
OY 577 TTGGCTCAGACTTTAAAGGACTTGTACCAATCCCATCTTTTCTGAACCTTTTATTC 636
OY 361 ggtgaagatactgaacttatttctaacttgaagaagacacacacacacacacacacac 420
DB GGTGAAGATTTGACATTTATTTTAACTTGAAAGGACCTTTCACAGAACCTTATTC 696
OY 637 GGTGAAGATTTGACATTTATTTTAACTTGAAAGGACCTTTCACAGAACCTTATTC 696
OY 421 aggaagaac 480
DB AGGAAGACCAACACCATCACCGCTGAGACACTGACCTTGAAATGGCCCTCCGCGC 756
OY 697 AGGAAGACCAACACCATCACCGCTGAGACACTGACCTTGAAATGGCCCTCCGCGC 756
OY 481 ctccaagacccctgac 540
DB CTCCAAGACCCCTGCATCATTTGAAGGAAATTCGCAAAAGGCAAGTCCACTCTGCT 816
OY 757 CTCCAAGACCCCTGCATCATTTGAAGGAAATTCGCAAAAGGCAAGTCCACTCTGCT 816
OY 541 cgaattgccaatgctcggagctcggagaagtgcaaggtctgaacaaagtccaattcgtc 600
DB CGCATTTGCCATGCTCTGGGCTCCGGAAGTGCNAAGGCTGTGACCAAGTTCAAAATTCGTC 876
OY 817 CGCATTTGCCATGCTCTGGGCTCCGGAAGTGCNAAGGCTGTGACCAAGTTCAAAATTCGTC 876
OY 601 tcttctcctcgtctcagcaagggcccaaggttgaaacttltgaaacctctgtgtacaac 660
DB TCTTCTCCTCGTCTCAGCAGGGCCCGAGGGTGACTTTTGAACCCCTCTGTGATCACTC 936
OY 877 TCTTCTCCTCGTCTCAGCAGGGCCCGAGGGTGACTTTTGAACCCCTCTGTGATCACTC 936
OY 661 ctggataatactggaacaatcagaagaagacacacacacacacacacacacacacacac 720
DB CTGGATAATCTTGACCAATCAGGAAGCAGACATTCATGCTCATGCTGGAAGCT 936
OY 937 CTGGATAATCTTGACCAATCAGGAAGCAGACATTCATGCTCATGCTGGAAGCT 936
OY 721 cagaaggtcttcttctccttctgtatggctacaatgaattcaagccccaagactccc 780
DB CAGAGGGTCTTCTTCTTCTTGTGATGCTCAATGAATTCAGCCCGCAGAACTCCCGC 1056
OY 997 CAGAGGGTCTTCTTCTTCTTGTGATGCTCAATGAATTCAGCCCGCAGAACTCCCGC 1056
OY 781 atcgaagccctgataaagaagaacacacacacacacacacacacacacacacacacac 840
DB ATCGAAGCCCTGATTAAGGAACCAACCGCTTCAAGAACATGCTCATGCTCACACATAC 1116
OY 1057 ATCGAAGCCCTGATTAAGGAACCAACCGCTTCAAGAACATGCTCATGCTCACACATAC 1116
OY 841 actgaatgctcgaagac 900
DB ACTGAATGCTCAGGACATACGCGCATTTGGTCCCTGACCTCTGAGGTGGGGAATG 1176
OY 1117 ACTGAATGCTCAGGACATACGCGCATTTGGTCCCTGACCTCTGAGGTGGGGAATG 1176
OY 901 acagaagaacagcccaagctctcatccagaagaagtgtcatcaagaagcttgctga 960
DB ACAGAAGAAGCCCGCAGGCTCTCATCTCCAGAAAGTGTGATCAAGAGCTTGTCTGA 1236
1177 ACAGAAGAAGCCCGCAGGCTCTCATCTCCAGAAAGTGTGATCAAGAGCTTGTCTGA 1236

OY 961 ttgtctcccaaatccagaatcccaagtgcttgaagaatctcaatgaagaacctctctt 1020
DB TTGTCTCCCAAAATCCAGAAATCCAGTGTCTTGAAGAAATCTCATGAAGAACCCCTCTCTT 1296
OY 1021 gttgcaatacactgtgcaatccagatgggtgaagaagtgaagttccacactccacacaa 1080
DB GTTGCAATACCTTGCAATCCAGATGGGGAAGGTGAGGTTCACCTCTCCACACAAACA 1356
OY 1297 GTTGCAATACCTTGCAATCCAGATGGGGAAGGTGAGGTTCACCTCTCCACACAAACA 1356
OY 1081 agcgttccataacctctcctatgactcgttgaatcagaagaagaacacacacacacacac 1140
DB AGCGTTCATACCTTCTATGATCTGTTGATACAGAAAACCAACCAATCAATTAAGGT 1416
OY 1357 AGCGTTCATACCTTCTATGATCTGTTGATACAGAAAACCAACCAATCAATTAAGGT 1416
OY 1141 gttgctggaagtgaattcaatcctggagcctggagcacttgaagaagactgaaggt 1200
DB GTTGCTGCAAGTACCTTCATTCGAGCCTTGACACCGTGGAACTACTGAGAGGT 1476
OY 1417 GTTGCTGCAAGTACCTTCATTCGAGCCTTGACACCGTGGAACTACTGAGAGGT 1476
OY 1201 gtttctcccaaatgttgaatttcgaacgcaagatgtgtccagcgtgaatgaatgtc 1260
DB GTTCTCCCAAGTGTGATTTGATTTGAGATGTCAGAGATGTCTCAGCGTGAATGAGATGTC 1536
OY 1477 GTTCTCCCAAGTGTGATTTGATTTGAGATGTCAGAGATGTCTCAGCGTGAATGAGATGTC 1536
OY 1261 gttgctgaacactgtggtcctctgttaatatagaagctcaaaagttccaagcaagtaaa 1320
DB GTTGCTGAACACTGGCTCTCTGTAAATATACGTCGAAAGGTTCAAGCCAAAGTATTA 1596
OY 1537 GTTGCTGAACACTGGCTCTCTGTAAATATACGTCGAAAGGTTCAAGCCAAAGTATTA 1596
OY 1321 tcttccaagaatcattccaagaagtaacaagaagcaagaagactcaagcttattgaag 1380
DB TCTTTCACAAAGTCATTCAGAGATACAGAGACAGAGGAACTCAGACTTCAGATTATGAGC 1656
OY 1597 TCTTTCACAAAGTCATTCAGAGATACAGAGACAGAGGAACTCAGACTTCAGATTATGAGC 1656
OY 1381 tctcaagaagcagaagagtgtagcacaaggggaatgtgttaacttgaagaagaatgttccat 1440
DB TCTCATGAGCCAGAGAGAGGACCAAGGGGAATGTACTTGCAGAAAATGTGTTCCATT 1716
OY 1657 TCTCATGAGCCAGAGAGAGGACCAAGGGGAATGTACTTGCAGAAAATGTGTTCCATT 1716
OY 1441 tctgaacttacaatcacttgaacacacacacacacacacacacacacacacacacacac 1500
DB TCGACATTAATCACTCACTTATATACACCTGCTCCGTAACCTGTCGTCATCTGTGAA 1776
OY 1717 TCGACATTAATCACTCACTTATATACACCTGCTCCGTAACCTGTCGTCATCTGTGAA 1776
OY 1501 gtcacagaagctgttgaagaagcctcgaagcagtgatcaagaagcctgctctcga 1560
DB GTCACAGAGGCTTTTGAAGCACTCGCAGACAGTATATCAACACGCTGCTCCGGA 1836
OY 1777 GTCACAGAGGCTTTTGAAGCACTCGCAGACAGTATATCAACACGCTGCTCCGGA 1836
OY 1561 ctccaatccgaagaagcctctctgaagaagaaacttltgaagaagtgtgaagaacacac 1620
DB CTCCAATCCGAAGGCTCTCTGAGACAGAAATCTTGTGAAGTGTGAAGAACAC 1896
OY 1837 CTCCAATCCGAAGGCTCTCTGAGACAGAAATCTTGTGAAGTGTGAAGAACAC 1896
OY 1621 actgaagcaagaatcttgaagaacataacaatcattccttgaagatgtgcatcact 1680
DB ACTGAGCAAGAAATCTTGAAAGCAATCAATCAATCTTGTGAGAGTGTGATCAT 1956
OY 1897 ACTGAGCAAGAAATCTTGAAAGCAATCAATCAATCTTGTGAGAGTGTGATCAT 1956
OY 1681 tatalcaagaagatacatccaatcagaacctgaagcagaagaattgaagcttcttcaa 1740
DB TATATCAAGAGATGATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 2016
OY 1957 TATATCAAGAGATGATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 2016
OY 1741 ggtaaaagcttatalcataactcagaagaacatcccgatctacttacttctgtga 1800
DB GTTAAAGCTTATATATCAATCAGGAACATCCCGATTAATTAATTAATTAATTAATTA 2076
OY 2017 GTTAAAGCTTATATATCAATCAGGAACATCCCGATTAATTAATTAATTAATTAATTA 2076
OY 1801 catltgccaatgttgaagtgctcgtgaactcattlaaactgaacttlataggggaagct 1860
DB CATTTGCCCAATGTGCAAGTGCCTGGAATTCATTAACCTGGAACCTTTATGCGGAGCT 2136
OY 2077 CATTTGCCCAATGTGCAAGTGCCTGGAATTCATTAACCTGGAACCTTTATGCGGAGCT 2136
OY 1861 atgagctcagtggaagaagcttgaagaagaacagagtggaatccaatgaagaagaagccca 1920
DB ATGAGCTTCAATGGGAAAAGCTGCAGAAACACAGAGTGCAGATCCACATTCGAAGAGCCCA 2196
OY 2137 ATGAGCTTCAATGGGAAAAGCTGCAGAAACACAGAGTGCAGATTCGAAGAGCCCA 2196
OY 1921 gaaactacaatcccaagcaggtgtatcttgttcttcaacttgaagaagaagaatccag 1980
DB GAAACTACATTCACAGCAGGCTGTATCTTGTCTTCAACTGGAAGCAGGATCAAG 2256
OY 2197 GAAACTACATTCACAGCAGGCTGTATCTTGTCTTCAACTGGAAGCAGGATCAAG 2256
OY 1981 actctggaagtgcaactccgggagttcaagaagaatgtgaataagaagatacatatctg 2040
DB ACTCTGAGAGTCACTCCGGGATTTTCAGCAAGTTAAATTAAGCAAGATATCAGATATCTG 2316
2257 ACTCTGAGAGTCACTCCGGGATTTTCAGCAAGTTAAATTAAGCAAGATATCAGATATCTG 2316

```

Oy 2041 gggaaataatcagctcctcccaagaagcctcagctcgaataaagaatgctgctg 2100
    |||||||
Db 2317 GGGAAATAATTCAGCTCTGCCCAAGCCTCAGGCTGCAATAAAGATGTCTGTGTG 2376
Oy 2101 gctggaagcctcagcttctgctcccgacacgctgtaagaacatttctctcagtg 2160
    |||||||
Db 2377 GCTGGAACCCACAGTTGGTCTCTCAGCACACCTGTAGAACAATTATTCTCTCATG 2436
Oy 2161 gccagtcctccctcacacatagaagatgagagacatcaatctgttacaacactgaa 2220
    |||||||
Db 2437 GCCAGTCCCTCACCATAGAGATGAGAGCACATCATCTGTAAACAACCTGAAACC 2496
Oy 2221 ttgaatcctcagctcagaatcaacagcgtccggtggtctcagcagacgcttggt 2280
    |||||||
Db 2497 TTGAGTATTTCATGACCTTACAGAAATTAACGGCTGCCGGTGTCTGACGACGCT 2556
Oy 2281 aacttgaagaacctacacagctcatalaagtacaataaagaatgaaatgaaatgct 2340
    |||||||
Db 2557 AACTTGAAGAACCCTTACAAAGCTCATATGATACATTAAGATGAATGAAGAAGATGCT 2616
Oy 2341 ataaactagctgaaagcctcgaataaagcctgaaagaagatgcttatttcaattg 2400
    |||||||
Db 2617 ATAAACAAGCTGAAAGGCGCTGAAAAACCTGAAGAAGATGTGTATTTCATTGACC 2676
Oy 2401 ttgtctgacatttgagagaggaatggaattacatagctcaagctcctgcaagtga 2460
    |||||||
Db 2677 TTGCTGTGACATTGGAGAGGAGGAAATGATTAACATAGTCAAGTCTCTGTCAAGT 2736
Oy 2461 gaccttgaagaatcaatctagctcctcctgctgctgctgctgcaaatgcaatg 2520
    |||||||
Db 2737 GACCTTGAAGAATTCATTAATGTCGTCGTCGCTTGTCTGCAAAATGCAATGCAAT 2796
Oy 2521 gctcagaatcttcacaaattgctcgaactgagcattcttgattatcagaataa 2580
    |||||||
Db 2797 GCTCAGAACTTCACAAATTTGGTCAAACTGAGCAATCTTGTATTATCAAAATTA 2856
Oy 2581 gaaaagaatggaatgaagcctcttcaatgaactgacagagatgaagctgtagaa 2640
    |||||||
Db 2857 GAAAAAGATGGAATGAAGCTCTTATATCACTGATCGACAGATGAAGCTGCTAGAA 2916
Oy 2641 ctcaacgacatgactgctcctcctggtggtgctgctgcaagcagcagcctgct 2700
    |||||||
Db 2917 CTCACCGCAGATGATGCTCCCTGGGGGCTGTGACGTGCAAGGAGCGCTGAGCA 2976
Oy 2701 aaacatttgagagaggtcccaacactgctcaagcttggttggttgaaactgag 2760
    |||||||
Db 2977 AAACATTTGGAGGAGCTCCCAACCTCTCAAGCTTGGGTTGAAAAAAGCTGAGAC 3036
Oy 2761 gatacagaatgataaattttagtgatcttcttgaaagaaccccttgaataaactc 2820
    |||||||
Db 3037 GATACAGAGATTAGAAATTTAGGTGCATTTTGTGAAAGAACCTCTGAAAAAAGCT 3096
Oy 2821 cagttgaaatctggcgggaacatgctgctgagcagtgatgagctgctcctcat 2880
    |||||||
Db 3097 CAGTTGAATTTGGCGGGAATCGTGTGAGCAGTATGATGGCTTGCCCTTCATGGGT 3156
Oy 2881 tttaggaatcttaagaactgctgttcttgacttagtaactaaagaattcttact 2940
    |||||||
Db 3157 TTTGGAATCTTAACCAATTAAGTGTGTGTGTTTGTGACTTAAAGAAATTTCTAC 3216
Oy 2941 ccagatgaatgcagaagaacttagccaagtgtatccaagttaacttctcagaag 3000
    |||||||
Db 3217 CCAGATTAATGTCAGAAAAGCTTAAGCAAGTGTATTCAGATTAACTTTCTGCAAA 3276
Oy 3001 aggcctgtctgggtgcaatttgatgataatgataatcagctgcttatacaggt 3060
    |||||||
Db 3277 AGGCTTGTGGGTGCAATTTGATGATGATGATGATGATGATGATGATGATGATG 3336
Oy 3061 ctatgaactgct 3072
    |||||||
Db 3337 CTAGTAAGTCTGCT 3348

```

```

RESULT 4
LOCUS AF376061
DEFINITION Homo sapiens caspase recruitment domain protein 12 mRNA, complete cds.
ACCESSION AF376061
VERSION AF376061.1
KEYWORDS GI:14040074
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3581)
AUTHORS Gingras, M., Qiu, J. and Margolin, J.F.
TITLE Differential expression of the caspase recruitment domain protein 12 (CARD12) during monocytic differentiation
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3581)
AUTHORS Gingras, M., Qiu, J. and Margolin, J.F.
TITLE Direct Submission
JOURNAL Submitted (03-MAY-2001) Pediatric/Texas Children's Cancer Center, Baylor College of Medicine, 6621 Fannin St., MC3-3320, Houston, TX 77030, USA
FEATURES
source location/Qualifiers
1. 3581
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2p21-22"
/cell_line="U937"
/cell_type="peripheral blood-derived CD14 mature monocyte"
490..3564
/note="CARD12"
/codon_start=1
/product="caspase recruitment domain protein 12"
/protein_id="AAK53443.1"
/db_xref="GI:14040075"
/translation="MNFIRKNSRALIORMKNTVYIKQTTDLFPVNVNLNREVVITCE
KVEDDANRGIITHMLKRGSESCNLELSLEKEMTPYLPYDINGSLRPHQTSGLDDIA
QDLKDYHTPSFLNFYPLGEDIDITENLKTSTFETPRVLMRDHNRHREQLKGLQA
LOSPLCIEGSGGKSTLLORIAMLMGSCCKALTKFKEVFLRLSAGQGLETLCD
QLDIPETIKOFEMALIKLROVLELDGYNFKRONCPETALIKHREKRNMYI
VTTEELRHIREGALTAEVGMTDSNALIREVILKLAGLILLOKLSQSLNLT
MKTPPLVYITCALQMGSESEPHSHQTTLFTFPDLQKKKKHKKYAAASDPTRSIDH
CGDLALGVFSHKFDELQVSVNEDALLTTLGLCKYTAQRKPKRPFHKSFOET
AGRLSSLSLSHEPEVTNGNGYLOKWSISDTSTYSLLRTCSVSATPAVMKH
LAAVYOHGCLGLSIAKRPMLROESLOSVSNTIEOELIKAININSEVEGILHYOEST
SKSALISOEPFAPFOGKSLYNSGNIPIYLFDFEHLPCASALDFIKIDYGGAMSW
EKAEDMGTHMEAPETYPISRAVSLFFMKQGEFTLFTYLDPSKLNKODIRYLKX
IFSSATSLRQIKRCAGVAGSLSVLSTCKNITSMEASPLTIEDBRHTSVTNLKT
LSLHLDQNLPGGLDSDLSGLNLTNLTKLINDNTKMEEDIKLAEGLKILKRCFLHL
THLSDIEGMDYIVKLSSEPCDLEELVSCCLSNVAVKILAEGLKILKRCFLHL
ENLEKNGENALHELIDRNMLPOLRLALMLPCGDOVGSLSLKLHEEPOLVKGL
KNRRLDTEIRIRIGAFEGKPNLNFPOOLNAGRVSDGMLAMGVENLKKLVFDFE
STKEFLPDPALVYKRLSGVLSKLFLDGEARLVGQFDDDLSTVTGFAKLVTA"
502..741
/note="Region: caspase recruitment domain"
BASE COUNT 1033 a 781 c 843 g 924 t
ORIGIN
Query Match 99.8%; Score 3065.6; DB 9; Length 3581;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 3068; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Oy 1 atgaattcataaagaacatagccgagccttctcaagaagaaggaaatgactgtata 60
    |||||||
Db 490 ATGAATTTCAATGAAGAGCAATAGCCGAGCCTTATTCAAGAAATGCGAATGACTTTATA 549
    |||||||
Oy 61 aagcaaatcacagatgacatttctgataatgctgtaacgcgagaagaatcaacatc 120
    |||||||
Db 550 AAGCAAAATCACAGATGACCTATTGTATGCAATCTTGTGAAATCCGAAGAAGTAACATC 609
    |||||||

```

QY 121 attgtctgcagaaagtgagagagatgctgtctagaagagatcatctatgattt:gaaa 180
Db 610 ATTGCTGGAGAGAGGTGGAGAGATGCTGCTAGAGAGATCATTCATGATTTT GCAAA 669
QY 181 aagaggtcagaagtcctgttaacctcttcttaaatcccttaagaagtggaacttc:tccta 240
Db 670 AAGGCTTCAAGCTCTGTAACTCTTTCTTAATTCCTTTAAGAGTGAACCTATC:TCCTA 729
QY 241 ttccagaacttgaaatgagcaagcttttctcatcagacatcagaagaagacttgg:cgat 300
Db 730 TTTCAGAGACTTGATGACAAAGCTTTTTCATCAGACATCAAGAGAGCTGG:CGAT 789
QY 301 ttggtcagaatttaaagaactgttaaccatacccaacttcttctgaacttlaa:ccct 360
Db 790 TTGGCTCAGATTAAAGAGCTTGTACCATATCCCATCTTTCTGAACTTTTATC:CCCT 849
QY 361 ggtgaagataatgacattatcttaacttgaagaagcaaccttcagaaactgtgc:gtg 420
Db 850 GGTGAAGATATTGACATTTATTTTAACTTGAAGACCTTTCACAGAACCACTCC:GTGG 909
QY 421 aggaaggaacaacaacatcacccgctgagagcagctgaacctgaatggcctcgc:ggct 480
Db 910 AGGAAGGACCAACACATCACCGCTGGAGACACTGACCTGAAATGGCCCTGCT:GGCT 969
QY 481 ctccagaagccctgcacatctagaaggaatctggcaagaagctccactctgc:ggag 540
Db 970 CTTCAGAGCCCTGTCATCATTTGAAGGGGATCTGGCAAGGCAAGTTCACCTGTG:GCAG 1029
QY 541 cgcattggcactgtcttggagctcgcgaagaatgcaagaacttgaaccaagtccaat:cgct 600
Db 1030 CGGATTGGCCATGCTCTGGGGCTCCGGAAATGCAAGGCTTGACCAAGTTCAAAATTC:GCTC 1089
QY 601 ttcttcctcgtctcagcaagggcccgaggtgtgacttttgaaaacctctgtatc:aatc 660
Db 1090 TTCTCTCCGCTCTCAGACAGGCGCCAGGTGGACTTTTGAACCCCTCTGTGATCACTC 1149
QY 661 ctgatatatactcgggacaaatcaagaagaacatctcatgagcaatgtctgaagct:tcgg 720
Db 1150 CTGCAATATACCTGACAAATCAGAGAACACATTCATGGCCATGCTGCGAAGCT:CCGG 1209
QY 721 cagaaggtctcttcttcttcttgaatgacaaatgaaatcaagcccaagactgcc:agaa 780
Db 1210 CAGAGGCTTCTTTCTCTTCTTGTGATGCTACAAATGAATCAAGCCCAAGAACTCC:CGAA 1269
QY 781 atcgaagccctgaataagaanaaacacccgcttcaagaacaatgctcatcgtc:caaac 840
Db 1270 ATCGAAGCCCTGTATAAAGGAACCAACCGCTTCAGAACATGCTCATCGCACAC:TAAC 1329
QY 841 actgaatgctcctgaggaacatacagcgagttgtgtccctgactgagtgagtgga:atg 900
Db 1330 ACTGAGTGCCTGAGGACATACGCGCATTTGGTGGCCCTGACTGAGGTGGGGGA:ATG 1389
QY 901 aacgaagaacagcgcccaaggtctcatccagaagtgctgaatcaagaaggtctgtga:agg 960
Db 1390 ACAGAAGACAGCCCGAGGCTCTCATCCGAGAGTGTGTATCAAGGAGGCTTGTGTA:AGCC 1449
QY 961 ttgttctcnaaatgaagaatccaggtgttgaagaatctcatgaagaccctct:ttt 1020
Db 1450 TTGTTGCTCAAAATTCAGAAATCCAGTGCTTGAAGAAATTCATGAAGAACCCCTCT:TTT 1309
QY 1021 gttgatacaactgtgcaatccagatgggtgaaagtgaatgtccaacttcacaaca:aca 1080
Db 1510 GTGGTATCATCTGTGCAATCCAGATGGGTGAAGTGAAGTTCCACTCCACACAA:ACA 1569
QY 1081 aagctgttccataactctctatgactgtgtgatacagaanaaacaaacaataa:agt 1140
Db 1570 ACCCTTTTCATACCTTCTTGTGATCTGTTGATACAGAAAAACAACCAACATTAAT:SGT 1629
QY 1141 gttgagcgaagtgaactcatctcgagaccttgaaacctgttgagagacctagctc:ggag 1200
Db 1630 GTGGCTGCAAGTGAATTCATTCGAGGCTTGACACACTGTGAGAACCTACGCTTGGA:GCT 1689
QY 1201 gtgtcttccacaagtgtgtcttgaactgcagaatgtgtccagcgtaagtggatgtc 1260

Db 1690 GTGTTCTCCACAAAGTTTGTATTTTCGAAGCTGCAGAGATGTGTCAGGCTGAATGAGATGGC 1749
QY 1261 ctgttcacaacttgagctccctctgttaatatagagctcaagaaggttcaagccaagtataa 1320
Db 1750 CTGCTGCAAACTGGGCTCCTCTTAATATACAGCTCAAAAGTTCAAGCCAAAGTATAA 1809
QY 1321 ttcttccaaagtcatctcagaaggtacacagcagaagaagactcaagcttataatgaag 1380
Db 1810 TTCTTTTCACAGTCAATTCAGGAGTACAGAGACAGAGAAAGTCAAGCTTATTGAGC 1869
QY 1381 tctcatgaagcagaagagtgagcaaggggaatgtgtacttgcagaanaatgtttccatt 1440
Db 1870 TCTCATGAGCCAGAGAGAGGTGACCAAGGGGAATGTTACTTGCAGAAATGTGTTCCATT 1929
QY 1441 tgggaattatcatccacttataagcaagctgtccggttaacactgtgggtcatctgtgaa 1500
Db 1930 TCGACATTAATCAATCCATTATACAGCTCTCCGGTACACCTGTGGTCACTGTGGAA 1989
QY 1501 gccaccaaggtgttataagacaccccgcaagcaggttatcaaacaggtctgcttcgga 1560
Db 1990 GCCACAGGGCTGTATGAAGCACCTCGCAGAGTGTATCAACACGCGCTGCTCGGA 2049
QY 1561 ctctcatcgcacaagaggtcctctctgagagaagaatctcttgcaaggtgtgaanaacacc 1620
Db 2050 CTTTCCATGCGCAAGAGGCTCTCTGAGACAGGATCTTTCGAAAGTGTGAANAACACC 2109
QY 1621 actgagcaagaataatctgaaagccataaacatcaatctcttctgtagaagtgtgacatc 1680
Db 2110 ACTGACCAAGAAATTTGTGAAGCCATTAACATCAATTCCTTGTGAAGTGTGCATCAT 2169
QY 1681 ttatatcaagagagatcatatccaaatcaagccctgagccaagaatttgaacttcttcaa 1740
Db 2170 TTATATCAAGAGTATCATTCCAATACGCCCTGAGCAAGAAATTTGAAGCTTCTTCA 2229
QY 1741 ggtlaaagcttatalatacaactcagaagaaatcccgattacttatttgaactctttaa 1800
Db 2230 GTTAAAGCTTATATATCAATCTCAGGAACATCCCGATTACTTATTTGAACTTTGAA 2289
QY 1801 catctgcccaatgtgcaaggtgtctgtgaacttcatataactggaacttataaggaggact 1860
Db 2290 CATTTCCCAATGTGTCAAGTGTCCCTGGACTTTCATTAACCTGACATTTATGGGGAGCT 2349
QY 1861 atggtctcatatgggaagaaggtctgcagaagaacacaggtgtgaatccacatggaagaaggccca 1920
Db 2350 ATGCTTCATATGGGAAAAGGCTGCAGAAAGCACAGGTGAATCCACATGGAAGAGGCCCA 2409
QY 1921 gaaacctatctcccaagcagaggtgtatcttcttcaactggaagcagaatccagg 1980
Db 2410 GAAACCTACATTCGCCAGAGGGCTGTATCTTGTCTTCAACTGGAAGCAGGAATTCAGG 2469
QY 1981 actctgaaggtcaactcccgagttccaagaagtgtaaatgaagaagatcatcatactcg 2040
Db 2470 ACTCTGAGGTACACCTCCGGGATTTACGAAGTTGAATGAAGCAATATCAATATTCG 2529
QY 2041 gggaaataatgaagctctgcacaaagcctcagagcttcaataaagaagatgtgtgtgtg 2100
Db 2530 GGGAAATATATCAAGCTCTGCCACAAGCCTTAGCTGCAAATAAGAGATGTGTGTGTG 2589
QY 2101 gctggaagctcagtttgttctcagacactgtgaagaacttatacttctcaagtgga 2160
Db 2590 GCTGGAAGCTCAGTTTGTGCTCAGACCTGTGAAGAACATTTATCTCATAGGTGGGA 2649
QY 2161 gccagtcctcccaacatagaagatggaagggacatcatctgtlaacaaacctgaaacc 2220
Db 2650 GCCAGTCCCTCCACATATGAGATGAGAGCACATCATCTGTAAACAAACCTGAAAAAC 2709
QY 2221 ttgaatlatcatgaactacagaaatcaacggctgcgggtgtgtctgtactgaagcttgggt 2280
Db 2710 TTGAGTATTCATGACCTACAGAAATCAACGGCTCCGGGTGTGTGACATCTTGAGT 2769
QY 2281 aacttgaagaaccttaagaagctcaatgaatgaataaagaatgaagaatgtgtc 2340


```
Db 2770 AACCTGGAAGACCTTACAAAGCTCATTAATGATTAACATTAAGATGAATGAAGATGCT 2829
Oy 2341 aataaactagctgaagcctgaagaaacctgaagaagatglttatcttaaccac 2400
Db 2830 ATTAACACTAGCTGAAGGCTTAACAACTGAAAGATGTTTATTTTCATTGACCCAC 2889
Oy 2401 ttgtctgacatttgagagaggaatgatatcacaatgacatctctgtcaagtgaacccgt 2460
Db 2890 TTGTCTGCATTTGAGAGGGAATGATTAACATGATCAAGTCTGTCAAGTGAACCCGT 2949
Oy 2461 gacctgaagaatccaattgactgtctgtctgtctgtctgtctgcaatgcagtgaaatcccta 2520
Db 2950 GACCTTGAAGAAATTCATATGATGCTCCGCTGCTTGTCTGCAAAATGACAGTGAATATCCTA 3009
Oy 2521 gctagaactctcaaatcttgtaaacctgagcaatcttctgtatctatcagaataattactg 2580
Db 3010 GCTCAGAACTTTCACAAATTTGTCAACTGACATCTTGTATTTATCAGAAATTTACCTG 3069
Oy 2581 gaaaaagatggaatgaagcctctcatgaactgacacgaagatgaacgtgctagaacag 2640
Db 3070 GAAAAAGATGAAATGAAGCTCTTCATGTAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 3129
Oy 2641 ctcaaccgcatgactgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 2700
Db 3130 CTCACCCGACTGATGCTCCCTGGGGGCTGTGACGTGCAAGCAGCAGCTGACGACCTGTTG 3189
Oy 2701 aaaaattggaggaggtcccaactcgtcaagcttggtgaaacctggaagactgacaca 2760
Db 3190 AAACATTTGGAGAGGAGTCCCAAACTGCTCAAGCTGGGTTGAAAACCTGAGACTCACA 3249
Oy 2761 gatacagagatcagaatctttagtgcattcttctgaaagaacctctgaaacctccag 2820
Db 3250 GATACAGAGATTTAGATTTTAGTGATTTTGTGAAAGAACCCCTGTAACCTGTAACCTGTAAC 3309
Oy 2821 cagttgaactctggcgagaaatcgtgtgagcagatgagatgagctgtgcttcaatggtgta 2880
Db 3310 CAGTTGATTTGGCGGGAATCGTGTGAGCAGTATGATGATGCTGCTGCTGCTGCTGCTGCTG 3369
Oy 2881 tttaagaactctaaagaattglttttttgaacttgaacttgaacttgaacttgaacttgaact 2940
Db 3370 TTTGAGATCTTAAGCAATTAATGTTTGTGACTTTAGTACTTAAGAAATTTCTACCTGAT 3429
Oy 2941 ccaagcatagtcagaaaactcagcaagtcgtatcacaagtaacttctcagaagact 3000
Db 3430 CCACCATTAAGTCAAAAACCTAGCCAAAGTGTATCCAAAGTATTTCTGCAAGAGACT 3489
Oy 3001 aggetctgtgggtgcaatttgatgatagtatcctcagtgltatcagaagtgctttaa 3060
Db 3490 AGGCTTGTGGGTGCAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3549
Oy 3061 ctatgaactgct 3072
Db 3550 CTAGTAAGTCT 3561
```

```
RESULT 5
AC010968
LOCUS AC010968 160583 bp DNA HTG 18-AUG-2000
DEFINITION Homo sapiens chromosome 2 clone RP11-9302, WORKING DRAFT SEQUENCE,
11 unordered pieces.
ACCESSION AC010968
VERSION AC010968.5 GI:9845170
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 160583)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 160583)
AUTHORS Waterston,R.H.
```

TITLE
JOURNAL
COMMENT

Direct Submission
Submitted (28-SEP-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Aug 18, 2000 this sequence version replaced gi:8439959.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
----- Project Information -----
Center project name: H_NH0093002
----- Summary Statistics -----
Sequencing vector: M13: 59%
Sequencing vector: plasmid: 41%
Chemistry: Dye-Primer ET: 47% of reads
Chemistry: Dye-terminator Big Dye: 53% of reads
Assembly program: Phrap: version 0.990319
Consensus quality: bases at least Q40
Consensus quality: bases at least Q30
Consensus quality: bases at least Q20
Insert size: 14700; agarose-fp
Insert size: 159583; sum-of-configs
Quality coverage: 6.64 in Q20 bases; agarose-fp
Quality coverage: 6.38 in Q20 bases; sum-of-configs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1080: contig of 1080 bp in length
* 1081 1180: gap of unknown length
* 1181 1476: contig of 1296 bp in length
* 2477 2576: gap of unknown length
* 2577 5035: contig of 2459 bp in length
* 5036 5135: gap of unknown length
* 5136 9606: contig of 4471 bp in length
* 9607 9706: gap of unknown length
* 9707 17078: contig of 7372 bp in length
* 17079 17178: gap of unknown length
* 17179 27158: contig of 9980 bp in length
* 27159 27258: gap of unknown length
* 27259 45137: contig of 17879 bp in length
* 45138 45237: gap of unknown length
* 45238 65522: contig of 20285 bp in length
* 65523 65623: gap of unknown length
* 65623 91489: contig of 25876 bp in length
* 91489 91598: gap of unknown length
* 91599 116835: contig of 25237 bp in length
* 116836 116935: gap of unknown length
* 116936 160583: contig of 43648 bp in length.
Location/Qualifiers
1. 160583
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/clone="RP11-9302"
1. 1080
/note="assembly_name:Contig9"
1181. 2476
/note="assembly_name:Contig12"
2577. 5035
/note="assembly_name:Contig13"
5136. 9606
/note="assembly_name:Contig14"
9707. 17078
/note="assembly_name:Contig15"
17179. 27158
/note="assembly_name:Contig16"

```
misc_feature 27259..45137
              /note="assembly_name:Contig17"
misc_feature 45238..65522
              /note="assembly_name:Contig18
              clone_end:17
              vector_side:right"
misc_feature 65623..91498
              /note="assembly_name:Contig19"
misc_feature 91599..116835
              /note="assembly_name:Contig20
              clone_end:SP6
              vector_side:right"
misc_feature 116936..160583
              /note="assembly_name:Contig21"
BASE COUNT 44349 a 33843 c 33703 g 47641 t 1047 others
ORIGIN
```

```
Query Match          65.2%  Score 2001.6:  DB 2:  Length 110583:
Best Local Similarity 97.4%  Pred No. 0:
Matches 2034; Conservative 0; Mismatches 54; Indels 0; Gaps 0;
```

```
261 aagctcttctacagacatcagaagagacttgacgatttgctcaagattcagaaga 320
Db 49186 AGCTCTTTTCACTACACACAGAGACTTGAGAGATTGGCTCAGGATTTTAAAG 49245
QY 321 ctgtacacacccacttctctgaacttacccttggtgaaatatatgaatcat 380
Db 49246 CTGTGTACCATACCCACTTTTCTGTGAACCTTTATCCCTTGAGAGATATTGACATTAT 49305
QY 381 tttaacttgaagaagacattccacagacactgtcctgtggaaggaagacacacatca 440
Db 49306 TTTTACTTTGAAAAGACACTTTCACAGAACCTGTCTGTGAGGAGACACACATCA 49365
QY 441 ccgcgtggaagcactacacctgaatgacctcctgcagagccttcacagaccctgcctcat 500
Db 49366 CCGCGTGGAGACACTGACCTGAAATGGCCGCGAGGCTCTTCAGAGCCCTGCGCTCAT 49425
QY 501 tgaagggagatcctggaaggaagcagctcactcctgctcagcagcagatgctcctgctc 560
Db 49426 TGAAGGGGAATCGCGAAAGGCAAGTCCACTCGTGCAGCCGATTTGCCATGCTCT 49485
QY 561 ctccggaagatggaagcctctgacagaatccaatcgaattccttctcctcctcctgcac 620
Db 49486 CTCGGGAAGATCGAAGGCTCTGACCAAGTTCAATTGCTCTTCTCCTCGCTGTCACAG 49545
QY 621 ggcgcaggtggaacttcttgaaacctcctgtgatacaactccctgagataacccctgacaat 680
Db 49546 GGCCTCAGGGGTGACTTTTGAAGCCCTGTGTATCAACTCTGTGATATACCTTGCAAAAT 49605
QY 681 cagaagaagacatcatgagcactgtctgtgaagctgcgcgacagaaggtctcttctcct 740
Db 49606 CAGGAAGCAACATTCATGTGCTGCTGAGAGCTGCGGCAAGAGGCTTCTTCTCTCT 49665
QY 741 tgaatgctacaatgaatctcaagccccaagaactgcccagaatatcgaaagccctgtat 800
Db 49666 TGATGCTCTCAATGAATTCAGGCCCAAGAACTGCCAGAAATGGAAGCCCTGTATTAAG 49725
QY 801 aaacccacccctcacaagaactgctcgtcactacactacactgaactgaactgcctgagc 860
Db 49726 AAACCAACCGCTTCAAGAACATGTGTCATCGTCACTACACTGAGTGGCTGAGGCTCAT 49785
QY 861 acggaagatttggtgcctcactacactgaagtggggaatagaagaagaagaagcccaagc 920
Db 49786 ACGGCAAGTTTGGTCCCTGACTGCTGAGGTGGGGAATATGACAGAAAGACAGCCCAAG 49845
QY 921 tctcactcagaagaatgctgatacaagaagcttgcctgaagagcttgctccaaattca 980
Db 49846 TCTCATCCGAGAAAGTCTGTATCAAGAGCTGTGTAAGGCTTGTCTCTCAAAATTTCA 49905
QY 981 atccaggtccttgaagaatctcaagaagccctcctcttgtgagtcacaaactgtgc 1040
Db 49906 ATCCAGGTCTTGAAGAAATCTCATGAAGACCCCTCTCTTGTGTGTCATCATCTTGTCT 49965
```

```
QY 1041 ccaagatgggtgaagtgagttcccaactctcaacacaaacacagctgttccatcctcta 1100
Db 49966 CCAGATGGGTGAAAGTGAAGTTCACACTCTCACACACAAACAGCTGTTCATACCTTTCA 50025
QY 1101 tgaatctgtgatacagaagaacaaacacacacaaataaaggtgtgagctgcaagtgactcat 1160
Db 50026 TGAATCTGTGATACAGAAAACAAACACAAACATTAAGGCTGTGCGTCGCAAGTCACTTCAAT 50085
QY 1161 tggagccctggaaccactgtgtgagacctcactcctggaagggtgtgtctctccacaagtctga 1220
Db 50086 TGGAGCCCTGGACCACTGTGGAGACCTACTCTGGAGGGTGTGTCTCCACAAAGTTTGA 50145
QY 1221 ttctggaactgcaagatgtgtccagcgtgaatgaagatgtcctgctgcaactgagctcct 1280
Db 50146 TTTGCAACTGCAAGATGTGTCCAGGCTGAATGAGATGTCTGTGACAACTGGGCTCTCT 50205
QY 1281 ctgtataataacagctcagaagttcaagccaaagratataatcttccaagatcacttcca 1340
Db 50206 CTGTAAATTTACAGCTCAAGGTTCAAGCCAAAGTAAATTTCTTTCACAACTCTTCCA 50265
QY 1341 ggaatcacacagaagaagagactcaagcagttatctgacgtctcatgagccagaagaggt 1400
Db 50266 GGAATTCACAGCAGAGAGAGACTCAGAGTATTTGAGCTGTCTCATGAGCCAGAGAGGT 50325
QY 1401 gaccaaagggaatggttactctgcaagaataatggttccattctggaatataccactta 1460
Db 50326 GACCAAGGGGAATGTTTACTTTCAGAAAATGCTTTCATTTGAGACTTATCAATCCACTTA 50385
QY 1461 tagcaagccctgctcgcgtacacactgtgtggtcactctgtgaagccacacagggctgtatga 1520
Db 50386 TAGCAAGCTGCTCCGCTGACACCTGTGTGTCATCTGTGAGGACACACCGCTGTATGTA 50445
QY 1521 gcaactcgcagaagtgatcaacaacagcctgctctcgcagacttccatctgcagaagagcc 1580
Db 50446 GCACTCTGCGAGCAGTGTATCAACACGCGTCCCTTTCGAGCTTTCATGCGCAAGAGGCC 50505
QY 1581 tctctggaagcaggaactctcttgaagaagtgtgaaaaaacacacactgagcaagaattctga 1640
Db 50506 TCTCTGGACACAGGAATCTTTGCCAAAGTGTGA AAAACACACTGAGCAAGAAATCTGTA 50565
QY 1641 agccataaatacaatctcttctgtgaagtgtgactcactatatacaagaagaagactc 1700
Db 50566 AGCCATAAATCAATCAATCTCTTGTAGAGTGTGCACTCTTTATATCAAGAGATGATC 50625
QY 1701 caaatcaagccctgagccaaagaatttgaagcttcttcaaggttaaaagcttatatacaa 1760
Db 50626 CAATATAGCCCTGAGCCAAAGAAATTTGAAAGCTTTTCAAGGTAAAGCTTATATATCA 50685
QY 1761 ctcaagggaacatcccgattacttatttgaacttcttgaacatttgcacattgtgcaag 1820
Db 50686 CTCAGGGAACATCCCGATTAATTTGATCTTGTGAACATTTGCCCAATTTGTGCAAG 50745
QY 1821 tgcctcgtgaactcaatlaaacttgaacttctataggggaagctatagcttcaatggaaaagc 1880
Db 50746 TGTCTGAGACTTCAATTAACCTGGACTTTTATGAGGAGCTATAGGCTTCAATGCGAAAAGC 50805
QY 1881 tgaagaagaacagaagtgagatccaacatgaagaagagcccaagaacatccatcccaagc 1940
Db 50806 TGCAGAGAAGACACAGGTGGAATCCACATGGAAGAGGCCCCAGAAACCTACATTCACAGCAG 50865
QY 1941 ggcctgatacttgccttcaactgaagcagaagaattcaagactctggaagtcaactcgc 2000
Db 50866 GGCCTGATCTTTTGTCTTCAACTGAGACAGAGATTCAGAGACTCTGAGGTCACTCCG 50925
QY 2001 ggaattcagaagaattgaataaagaagaatatacaataatcggggaataatataagctctgc 2060
Db 50926 GGAATTTCAAGAAATTTGAATGAAGAGATATCAATATATGAGGAAATATCAAGCTCTCC 50985
QY 2061 cacaagcctcaggtctgcaataaagaagatgtgctggtgtgctggaagccctcagtttgt 2120
Db 50986 CACAAGCCTCAGGCTGCAAAATTAAGAGATGTCTGTGTGCTGGAAGCTCAGATTGTGT 51045
```


QY	2115	tttggtctctgaaccccgctgaagaacttta	ttctctcattgctggtggaaagccagccccc	taac	2117
Db	361	TTTGGTCTCTGACGACCGGTGAAGACATTTATTTCTCTCATGTGTGAAGCCGACGTCCTCTAC			420
QY	2175	catagaagatgtagggagacataccacatctgttaacaaacc	tgaaaaacccttgatgattcattga		2234
Db	421	CATGAGAGATGAGAGGCGACATTCACATCTGTGAACAAACCTGAAACCTTGAGATTCATGA			480
QY	2235	ccctacgaatcaacagcgtctccgggtgggtctgactgacagcttgggtlaacttgaagaacct			2294
Db	481	CCCTACGAATCAACGGCGTCCGGGTGGTGTGACTGACACTTGCGGTAACCTTGAAACACCT			540
QY	2295	tacaaagctctaatatgatatcacataaagttgaat	tgaaagaaggtgtctataaactgctga		2354
Db	541	TACAAAGCTCTAATATGATATACATATAAGATGAATGAAGAAGATGCTATATAAATCTGACTGA			600
QY	2355	agagcctgaaaaaacctgtaagaaga	tgtgttatttcaattgaccaccactgtctgacatctgg		2414
Db	601	AGGCGTGA AAAACCTGAGAAAGATGTGTATTATTTCTATTGGACCACTGTCTGACATTTGG			660
QY	2415	agagggagatgatatcaatagtaagctctctgtcaag	tgaaccctgtgacttgaagaat		2474
Db	661	AGAGGGATGGATTTACTATGATCAAGTCTGTGTCAMGTGAACCTGTGACTCTGGAAGAAAT			720
QY	2475	tcaattgtctctgcgcgtctgtctc	taaatgaaagaaatccctagctccgaactctta		2534
Db	721	TCAATTAAGTCTCCGCGCTGTGTCTCAATAGAGGAAATCTAGCTCAACAATCTTCA			780
QY	2535	caatttgcataaactgagacttcttgat	attatataaagaatactaaccttgaaagaatgagaa		2594
Db	781	CAATTTGCTCAACTGACGATCTCTGATTTATCAGAAATTTACCTGSGAAAAAGATGGAA			840
QY	2595	tgaagctctctcaatgaaactgata	tgcagagatgaagctgtctagaaagcttcaaccgacatgat		2654
Db	841	TGAAGCTCTTTCATGAACATGTGATGCACAGGATGAACCTGTCTAGAAACGCTCACCGCACTGAT			900
QY	2655	gctgcctctggggcgtgtgacgt	tgcagaagcgacgttcgaagcatttgaagaat		2714
Db	901	GCTGCCCTGGGGCTGTGACCTGCCAAGGCAGCCTGACACAGCTGTGTGAACATTTGGAGGA			960
QY	2715	ggtcccaacaactcgtcaagctt	gtgttgaaaaaactgagaactcaagaatacagagattag		2774
Db	961	GGTCCCAACAACCTGCTCAAGCTTGGGTGA AAAACCTGGAAGTCAACAGATACGAGATTAG			1020
QY	2775	aattttagtgcattctt	ttggaagaaccctctgaaaaactccagcagttgaattggc		2834
Db	1021	AATTTTAGGGTGACATTTTGTGAAAAGAACCTCTGAAAACCTTCCAGCACTTGAAATTTGGC			1080
QY	2835	gggaaatcgtgttagcagtgata	tgaaatggccttgcctcatcaggtgtatattggaagaactctaa		2894
Db	1081	GGGAAATCGTGTAGCGAGGTATGTGATGTGGCTTGCTTCAATGGGTGTATTTGGAACATCTTAA			1140
QY	2895	gcaattagtgcttctt	gtacttagtaactaaagaatttctaacctgtatccagatctagtcag		2954
Db	1141	GCAATTTAGTGTTTTGTGACTTTAGTACTATAAGAAATTTTCACTGTATCCAGCATTTAGTGTAG			1200
QY	2955	aaaacttagccaagtgatataccaagttaact	ctttctgcagaagaagctagcgtgtgttggtg		3014
Db	1201	AAAACTTAGCCCAAGTGTATTCACAAGTTAACTTTTTCGCAAGAAGCTAGGGCTGTGGGGTG			1260
QY	3015	gcaacttgatgatgatgatc	atcctgaatttctaacagtgtcttcttaactagtaactgctc		3072
Db	1261	GCAATTTGATGATGATGATGATCTCAAGTGTATTTACAGGTGCTTTTAAACATGATACGTCT			1318
RESULT	8				
AY027788		AY027788	1395 bp	mrna	PRI
LOCUS		Homo sapiens	CLANB (CLANB)	mrna, complete cds.	20-JUL-2001
DEFINITION		AY027788			
ACCESSION		AY027788.1	GI:14324114		
VERSION					
KEYWORDS					

Source	Organism	Human
REFERENCE	human.	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
AUTHORS	1 (bases 1 to 1395)	
TITLE	Damiano,J.S., Stehlik,C., Plo,F., Godzik,A. and Reed,J.C.	
JOURNAL	Clan, a novel human ced-4-like gene	
MEDLINE	Genomics. 75 (1-3), 77-83 (2001)	
PUBMED	21365712	
REFERENCE	11472070	
AUTHORS	2 (bases 1 to 1395)	
TITLE	Stehlik,C., Damiano,J.S., Plo,F., Godzik,A. and Reed,J.C.	
JOURNAL	Direct Submission	
FEATURES	Submitted (21-FEB-2001) Program on Apoptosis and Cell Death Research, The Burnham Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA	
source	Location/Qualifiers	
gene	1..1395	
CDS	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/chromosome="2"	
	/map="2p22-p21"	
	/issue_type="Lung"	
	1..1395	
	/gene="CLAN1"	
	277..1356	
	/gene="CLAN1"	
	/codon_start=1	
	/product="CLAMB"	
	/protein_id="AAK14777.1"	
	/db_xref="GI:14324115"	
	/translation="MNFKDNRSRLIQRMGTVIKQITDDLFWNVNLRREVNICEE	
	KVEDADARGIHMLIKGSGESCNLFSLSEEMNPPLPQDLNGSGLDSIGNLNKLR	
	LIMDNIRNEDDAIKIIEGLKLNKMLPFLHLSIDIGEMDYIVKLSSEPDLEI	
	QVSCGSIANAIVKTLAONLHNVKSLTDLSENVLEKDNQAEHLIDRMNVLEQLTA	
	LMPLSGDVGSGSLSLKHLHEVYQVLYKLSLKNRRLDTETRIIGAFEGKNPLKNPO	
	LNLAGNVSSDGLAFMGVFNENLKQLVFFDFSTKEFLPDALVRKLSQVLSKTFLOE	
	ARLVGWPFDDDDLSVITGAERKLVTA"	
BASE COUNT	436 a 248 c 327 g 384 t	
ORIGIN		
Query Match	26.5%; Score 815.4; DB 9; Length 1395;	
Best Local Similarity	95.9%; Pred. No. 5.2e-208;	
Matches 837; Conservative	0; Mismatches 36; Indels 0; Gaps 0;	
QY	2200	tctgtgaacaacccggaacacttgtagtatcttaacgaccacgaatcaacgagctccgaggt 2259
Db	481	TTTTCTTAATCCCTTAAAGAGGTGAACATATCTCTATTTCAGAGACTTGGAATGCAAAAGT 540
QY	2260	ggtctgactgacagcttggtgaacttggaagaccttacaagctcataatgataacata 2319
Db	541	GGTCGACTGACAGCTTGCGTGAAGTGAAGAACCTTACAAAGCTCATATATGATRAACTTA 600
QY	2320	aagatgaatgaagaagatgctataaaactagactgaagcgctgaaanaacctgaagaagatg 2379
Db	601	AAGATGATGAAGAAGATGCTATATAAAGTACGTGAGGCGCTGAAAAACCTGAAGAAGATG 660
QY	2380	tgttatattcatcttgacccaacttgctctgaacatggagaagggaatggatatgaatgaag 2439
Db	661	TGTTATATTTCATTGACCCACTTGCTGCTCAATGTGAGAGGGAATGGAATTAATATGTCAG 720
QY	2440	tctctgcaagtgaaccccttgagaccttgaaagaatctcaatgaatcctcctgctgctgct 2499
Db	721	TCTCTGTCAAAGTGAAGCCTTGAGACCTTGAAGAATTCATTAATAGTCTCCTGCTGCTGCT 780
QY	2500	gcaaatgcagtgaaatctcctagctcagaactctcaacaattggtcaaacctgagcatctt 2559
Db	781	GCAATGCAAGTGAAGAATCTTAGCTCAGAAATCTTCAACATTTTGTCACAACTGACGATCTT 840
QY	2560	gattatcaagaatattaccttgagaaagaatggaaatgaagctcttcacatgaactgacatgac 2619
Db	841	GATTTATCAGAAATATTACCTTGAAAAAGATGAATGAAGCTCTTTCATGAACTGATGCAC 900

```

QY 2620 aggaataagctgctagacagctcaccgacgtatgctgcctggggctgagagcga 2679
    |||
DB 901 AGATGACAGCTGTAGACAGCTCAGCGCACTGATGCTGCGCTGGAGCTGACG:GCAA 960
QY 2680 ggcagcctgagcagcctgtctgaaacatttgagagaggtcccaacatcgltcaagcltggg 2739
    |||
DB 961 GGCAGCCTGAGCAGCCTGTTGAACATTTGAGAGAGGTCCACACACTGCTCAAGCTTGGG 1020
QY 2740 tggaaaacttgagagactcacagatcacagagattgaaatttgagtgatcttttg:aaag 2799
    |||
DB 1021 TTGAAAAACGGAGACACAGATACAGATTAAGATTTAGCTGATTTTTC:AAAG 1080
QY 2800 aaccctctgaaaaactcaccgagcttgaaatttgcgggaaatcgltgagcagctg:tgga 2859
    |||
DB 1081 AACCCCTGAAAAAATCTCCAGCAGTTGAAATTTGCGGAAATCGTGAGCAGAGT:TGGA 1140
QY 2860 tggcttgcttcataggctgctatctgaaacattcaagaattagtgcttttgaacttga 2919
    |||
DB 1141 TGGCTTCCTTCATGCGTGTATTGAGATCTTACCAATTTAGTGTTTTGACTT:TAGT 1200
QY 2920 actaaaagaattcttaacttgatccagcattagtcagaaaacttagccaaagtgtatc:caag 2979
    |||
DB 1201 ACTAAAGAAATTTCTACTGATCCAGCATTTAGTCAGAAACTTAGCCAAAGTTTATC:CAAG 1260
QY 2980 ttaacttctgcaagaagagctgagctgtgtggtggcaatttgaatgagatgctcagt 3039
    |||
DB 1261 TTAACTTTTGCAGAGAGCTAGGCTGTGGTGGCAATTTGATGATGATGATCTCAGT 1320
QY 3040 gttattacaagtgctttaaactgatactgct 3072
    |||
DB 1321 GTTATTACAGCTGCTTTTAACTAGTACTGCT 1353

```

```

RESULT 9
LOCUS AY027789 768 bp mRNA PRI 20-JUL-2001
DEFINITION Homo sapiens CLANC (CLANI) mRNA, complete cds.
ACCESSION AY027789
VERSION AY027789.1 GI:14324116
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 768)
AUTHORS Damiano,J.S., Stehlik,C., Pio,F., Godzik,A. and Reed,J.C.
JOURNAL JOURNAL OF CLINICAL GENOMICS 75 (1-3), 77-83 (2001)
MEDLINE 21365712
PUBMED 11472070
2 (bases 1 to 768)
REFERENCE 2 (bases 1 to 768)
AUTHORS Stehlik,C., Damiano,J.S., Pio,F., Godzik,A. and Reed,J.C.
JOURNAL Direct Submission
TITLE Submitted (21-FEB-2001) Program on Apoptosis and Cell Death
Research, The Burnham Institute, 10901 North Torrey Pines Road, La
Jolla, CA 92037, USA
FEATURES
source
1..768
location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2p22-p21"
/lssue-type="lung"
1..768
/gene="CLANI"
277..747
/gene="CLANI"
/codon_start=1
/product="CLANC"
/protein_id="AAK14778.1"
/db_xref="GI:14324117"
/translation="MNFIDKNSRALIQRMGMTVVKIQITDDELFWVN/LNREVNITICE

```

```

XVEQDAARGLTHMLKKGSESCNLFKSKIKENNYPLFQDLNQSILFQHSSEGLDILA
QDLKDIHTFSPSLNFTPLGEIDIDILNLSSTFETPLMKDRHHRVEQLTVL"
BASE COUNT 218 a 157 c 180 g 213 t
ORIGIN
Query Match 15.0%; Score 461; DB 9; Length 768;
Best Local Similarity 100.0%; Pred. No. 9,7e-113;
Matches 461; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 atgaattctaaaggacaaatagccgagcccttattcaaaagatgggaatgctgtata 60
    |||
DB 277 ATGAATTTCTAAAGGACAAATFACCCGAGCCCTTATTCAAAGAAATGGAAATGACTGTATA 336
QY 61 aagcaaatcacagatgacctattgtatgaaatgcttcgacgcaagaagtaaacatc 120
    |||
DB 337 AACCAATFACAGATGACCTATTGTGTGAAATGTTCTGATCGCGAAGTAACATC 396
QY 121 attgctcgcaagaagtgagcagagatgctgctagagagatcattcaatgatttggaa 180
    |||
DB 397 ATTTGCTGCGAGAAAGTGAGAGCATGCTGTAGAGGATCATTCACATGATTTTGAAA 456
QY 181 aagggttcaagatcctgtaaccccttcttaaatcccttaagaagtggaacttctcta 240
    |||
DB 457 AAGGTTCAAGTCCTTAACCTCTTTCTTAATCCCTTAAGAGATGGAACATCTCTCTA 516
QY 241 ttcaagacttgaatggacaaagtcttttcatcagacatcagaagaagacttgaagat 300
    |||
DB 517 TTTCAGACTTGTAATGACAAAGCTTTTTCATCAGACATCAAGAGACTGAGAGAT 576
QY 301 ttgctcaggaatttaagaagctgtacataaccacatcttcttgaaactttatccctt 360
    |||
DB 577 TTGGCTCAGATTTTAAAGACATTCACATACCCCATCTTGTGACTTTATCCCTT 636
QY 361 ggtgaagatatgacattattttaaacttgaagaagaccttcaagaacctgtctgtgg 420
    |||
DB 637 GGTGAAGATTTTGAATATTTTAACTTGAAGAACCTTCAAGAACTGCTGCTGG 696
QY 421 aggaagagaccaaaccatcccgctgagagcagctgacct 461
    |||
DB 697 AGGAAGGACCAACACCATCACCGCTGAGACACCTGACCT 737
RESULT 10
LOCUS AC011232 185281 bp DNA HTG 10-MAR-2001
DEFINITION Homo sapiens chromosome 2 clone RP11-78E13, WORKING DRAFT SEQUENCE,
7 unordered pieces.
ACCESSION AC011232
VERSION AC011232.7 GI:13270720
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 185281)
AUTHORS Waterston,R.H.
JOURNAL The sequence of Homo sapiens clone
TITLE Unpublished
AUTHORS Waterston,R.H.
JOURNAL Direct Submission
TITLE Submitted (04-OCT-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT
On Mar 10, 2001 this sequence version replaced gi:9799811.
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:http://genome.wustl.edu/gsc/index.shtml
Project Information -----
Center project name: H_NH0078E13

```

Summary Statistics

Sequencing vector: M13; 578
 Chemistry: Dye-terminator Big Dye; 50% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 182798 bases at least Q40
 Consensus quality: 183538 bases at least Q30
 Consensus quality: 184045 bases at least Q20
 Insert size: 168000; agarose-fp
 Insert size: 184681; sum-of-contigs
 Quality coverage: 7.07 in Q20 bases; agarose-fp
 Quality coverage: 7.33 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 7 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1985: contig of 1985 bp in length
 1986 2085: gap of unknown length
 2086 8183: contig of 6098 bp in length
 8184 8283: gap of unknown length
 8284 22741: contig of 14458 bp in length
 22742 22841: gap of unknown length
 22842 41446: contig of 18605 bp in length
 41447 41546: gap of unknown length
 41547 69182: contig of 27636 bp in length
 69183 69282: gap of unknown length
 69283 112131: contig of 42849 bp in length
 112132 112231: gap of unknown length
 112232 185281: contig of 73050 bp in length.
 185281 Location/Qualifiers

FEATURES

source
 1. 185281
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /clone="RP11-78E13"
 1. 1985
 /note="assembly_name:Contig11"
 2086. 8183
 /note="assembly_name:Contig12"
 8284. 22741
 /note="assembly_name:Contig13"
 22842. 41446
 /note="assembly_name:Contig14"
 41547. 69182
 /note="assembly_name:Contig15"
 69283. 112131
 /note="assembly_name:Contig16"
 112232. 185281
 /note="assembly_name:Contig17"
 BASE COUNT 52719 a 36957 c 37187 g 57816 t 602 others
 ORIGIN

Query Match 9.5%; Score 292.2; DB 2; Length 185281;
 Best Local Similarity 85.9%; Pred. No. 5.5e-67;
 Matches 324; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
 QY 1972 gaattcgaactctgaggtcacaactccggagattcgaagaattgaataagaagatc 2031
 |||
 Db 185281 GAATTCAGGACTCTGGAGGTCACACTCCGGGATTTCAGCAAGTTGAATACCAAGATATC 185222
 |||
 QY 2032 acatattcgggaaataatcagctctgcacaaagcctcaagctgcagcaataaagaagatg 2091
 |||
 Db 185221 AGATATCTGGGGAAATATTCAGCTCTGCCACAAGCCTCAAGCTGCATTAATAAGATGT 185162
 |||
 QY 2092 gctgtgtgctggaagcctcagcttgctcctcagcacctgtaagaacattatctctc 2151
 |||

Db 185161 GCTGTGTGCGCTGGAAGCCTCAGTTTGCTCCTCAGACACCTGTAGAACATTTATCTCTC 185102
 QY 2152 atggttgaagcagtcctccatcatagaatgagagacatcatctcgttaacaac 2211
 |||
 Db 185101 ATGTGTGAAGCCAGTCCCTCACCATTAGATGATGAGAGGACATCATCTCTGTAACAAAC 185042
 |||
 QY 2212 ctgaaaccttgatattcatgacctacagaatcaagagcgcgggtggtctactac 2271
 |||
 Db 185041 CTGAAACCTTGATATTTCATGACCTACAGATACAGCGCTCCGGGATTGTAATATTC 184982
 |||
 QY 2272 agcttggtaacttgaagaaccttacaagctcataatgataaataagatgaatgaa 2331
 |||
 Db 184981 AGTGTTGTGCTCTTGTGCTTAATAAAAAATAACAGTATATAATTTGGAAAAAGAGA 184922
 |||
 QY 2332 gaagatgctataaact 2348
 |||
 Db 184921 GAGGAGACTTATTTCT 184905
 |||

RESULT 11

CNS01DS8/c
 LOCUS
 DEFINITION
 BAC sequence from the SPG4 candidate region at 2p21-2p22 BAC 563N04
 of RPEC-11 library from chromosome 2 of Homo sapiens (Human),
 complete sequence.

ACCESSION
 AL121658 2 GI:7159619
 VERSION
 HTG; HTGS-DRAFT; SPG4 genomic DNA interval.
 KEYWORDS
 SOURCE
 ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 (bases 1 to 162692)
 Hazan,J., Fonkheche,N., Mavel,D., Paternotte,C., Samson,D.,
 Attiguenave,F., Davoine,C.S., Cruaud,C., Burr,A., Wincker,P.,
 Brothier,P., Catolico,L., Barbe,V., Burgunder,J.M.,
 Prud'Homme,J.F., Brice,A., Fontaine,B., Hellig,R. and
 Weissenbach,J.

TITLE
 Spastin, a novel AAA protein, is altered in the most frequent form
 of autosomal dominant spastic paraplegia
 JOURNAL
 Nat. Genet. (1999) In press
 REFERENCE
 2 (bases 1 to 162692)
 Genoscope.

JOURNAL
 Submitted (23-FEB-2000) to the EMBL/GenBank/DBJ databases
 COMMENT
 On Mar 6, 2000 this sequence version replaced gi:6002391.
 FEATURES
 Location/Qualifiers

1. 162692
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /clone="RP11-78E13"
 /clone="563N04"
 BASE COUNT 44981 a 32017 c 32887 g 52792 t 15 others
 ORIGIN

Query Match 9.5%; Score 292; DB 9; Length 162692;
 Best Local Similarity 100.0%; Pred. No. 6.1e-67;
 Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2781 aggtgcatcttttggaaagaccctcgaanaactccagcaagttgaatttgggggaa 2840
 |||
 Db 152218 AGGTGCAATTTTGGAAAGAACCTCTGAAAAACTCCAGCAAGTTGAATTTGGCGGGA 152159
 |||
 QY 2841 tctgttgcagcagatgatgctgctcctcatgtggttlatcttgaacttaagcaact 2900
 |||
 Db 152158 TCGTGTCAGCAGTATGATGATGGCTTGCTCATGGGTATATTAGATCTTAAGCAATT 152099
 |||
 QY 2901 agtgttttcttacttactaagaattctctactgataccagcattagtcagaact 2960
 |||
 Db 152098 AGTGTTTTTACTTATGACTAAGAAATTTCTACCTGATCAGCATAGTCAGAAAACT 152039
 |||

OY 2961 tagcgaagtgatcaccaggttcttcctcgaagaagcttagcgtgtgtggtggaatt 3020
|||||
Db 152038 TAGCCAAAGTGTATCCAGTTACTTTCTCGCAAGAGCTAGGCTTGCGGTGG:AAIT 151979
OY 3021 tgatgatgatgatcctcagtgatcctcaggtgtctttaaactcagtaactgct 3072
|||||
Db 151978 TGATGATGATGATCTCAGTGTTATTCAGGTGCTTTTAACTAGTACTGCT 151927

RESULT 12
AY027790 578 bp mRNA PRI 20-JUL-2001
LOCUS Homo sapiens CLAND (CLAN1) mRNA, complete cds.
DEFINITION
ACCESSION AY027790
VERSION AY027790.1 GI:14324118
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
1 (bases 1 to 578)
Damiano,J.S., Stehlik,C., Pio,F., Godzik,A. and Reed,J.C.
Clan, a novel human ced-4-like gene
Genomics. 75 (1-3), 77-83 (2001)
JOURNAL 21365712
MEDLINE 11472070
PUBMED
REFERENCE
2 (bases 1 to 578)
Stehlik,C., Damiano,J.S., Pio,F., Godzik,A. and Reed,J.C.
Direct Submission
Submitted (21-FEB-2001) Program on Apoptosis and Cell Death
Research The Burnham Institute, 10901 North Torrey Pines Road, La
Jolla, CA 92037, USA

FEATURES
source
1..578
location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2p22-p21"
/tissue_type="lung"
1..578
/gene="CLAN1"
277..555
/gene="CLAN1"
/note="CARD protein"
/codon_start=1
/product="CLAND"
/protein_id="AAK14779.1"
/db_xref="GI:14324119"
/translation="MNFKIDNSRALIQRMGNTVIKQITDDLFVNVVLRREVNITCE
KVEDDARGITHTMKKSGESCNEFLKSLKEMVNPFLFDLNGQSILTA"
BASE COUNT 172 a 106 c 143 g 157 t
ORIGIN

Query Match 8.8%; Score 269; DB 9; Length 578;
Best Local Similarity 100.0%; Pred. No. 4e-61;
Matches 269; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atgaattcataagaacaatagccgagccctattcaagaatgggaatgactgtcata 60
Db 277 ATGAATTTTCATTAAGGACATAGCCGAGCCCTTATTCAAGAATGGGAATGATGTAATA 336
OY 61 aagcaatcacagatgacctattgtatggaatgtctcgaatcggaagaaglaaa aatc 120
Db 337 AAGCAAAATCACAGATGACCTATTGTATGGAATGTTCTGAATCGGAAGAAGTAATAA:ATTC 396
OY 121 atttgctcgagaaggtggaagcagatgctgttaagaaggatcatcacaatgattttcaaa 180
Db 397 ATTGCTCGAGAAAGTGAGAGCATGCTCTAGAGGATCATTCACATGATTTT:AAAA 456
OY 181 aagggtcagagtcctcgaacctctcttcaatcccttaagaagatggaatcacc:cta 240
Db 457 AAGGTTCAAGATCCTGTAACCTCTTTCTTAATCCCTTAAGAGATGGAATATATCC:CTA 516

OY 241 ttccagacttgaatgacaaagctttt 269
Db 517 TTTCAGGACTTGAAATGACAAAGCTTTT 545

RESULT 13
G55568/c 553 bp DNA STS 30-MAR-2000
LOCUS SHGC-100923 Human Homo sapiens STS genomic, sequence tagged site.
DEFINITION
ACCESSION G55568
VERSION G55568.1 GI:6120887
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
1 (bases 1 to 553)
Olivier,M. and Cox,D.R.
Unpublished, Olivier, M., Cox, D.R. (2000)
JOURNAL Unpublished (2000)
COMMENT
Contact: Michael Olivier, David R. Cox
Stanford Human Genome Center
Stanford University School of Medicine
4005 Miranda Ave. 2nd Fl., Palo Alto, CA 94025, USA
Tel: (650) 320-5800
Fax: (650) 320-5801
Email: olivier@shgc.stanford.edu
Primer A: AATPAGSGGCAATATAGCAAA
Primer B: TACACCCCTTTGTTCTCTCA
STS size: 322

PCR Profile:
Initial incubation: 95 degrees C for 10 minutes
Denaturation: 94 degrees C for 30 seconds
Annealing: 60 degrees C for 30 seconds
Polymerization: 72 degrees C for 23 seconds
PCR Cycles: 30
Thermal Cycler: Perkin Elmer 9700
Template: 25 ng
Primer: each 1 uM
dNTPs: each 200 uM
Amplifrag Gold Polymerase: 0.07 units/ul
Total Vol: 5 ul

Buffer:
MgCl2: 2.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
pH: 8.3

BAC ends sequenced at TIGR from the RPC11 BAC library. Designed and developed at the Stanford Human Genome Center.

FEATURES
source
1..553
location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="2"
/clone_11b="Human"
STS
primer_bind 5..27
primer_bind complement(304..326)
BASE COUNT 170 a 107 c 114 g 162 t
ORIGIN

Query Match 5.5%; Score 170; DB 11; Length 553;
Best Local Similarity 100.0%; Pred. No. 1.7e-34;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 2615 tcacacagatgaagctgctcagacagctcaccgacatgctgcgcctgagctgacg 2674
|||||

	exon	/complement(79055..79174) /gene="Naip3" /note="region of sequence similarity; not known to be transcribed"	
	exon	/number=7 /complement(90835..90887) /gene="Naip3" /note="region of sequence similarity; not known to be transcribed"	
	exon	/number=6 /complement(91714..91794) /gene="Naip3" /note="region of sequence similarity; not known to be transcribed"	
	exon	/number=5 /complement(94865..94966) /gene="Naip3" /note="region of sequence similarity; not known to be transcribed"	
	exon	/number=4 /complement(95293..95983) /gene="Naip3" /note="region of sequence similarity; not known to be transcribed"	
	exon	/number=3 /complement(106079..106141) /gene="Naip3" /note="region of sequence similarity; not known to be transcribed"	
BASE COUNT	35695 a 24748 c 24777 g 32571 t		
ORIGIN			
Query Match	2-2% Score 68.4; DB 10; Length 117791.		
Best Local Similarity	53.0%; Pred. No. 7.8e-07;		
Matches	197; Conservative 0; Mismatches 166; Indels 9; Gaps 2.		
Oy	500 ttgaagggaatcttgcaagaaagcattccactctgtcgacgcgattgccai/gctctggg 559		
Db	17517 TGGAGGCGGACCTGGAAGTGTAAAGAACCAACTTCCTGAGAGGATAGCTTTTCTGTGGG 17458		
Oy	560 gctccggaagtgcgaagctctgaccagaattcaaattcgcttcttcctcg :ct----- 614		
Db	17457 CATCAGAGTGCCGCCCTGTTGTACAGGTTCCAAGTGTCTTCTTACCTCTTAGATT 17398		
Oy	615 -cagcagggcccaggglygaccttttgaaccctctgatactaactcctgataatactgy 673		
Db	17397 CCATCACACCATCAGGAGACAGGCACACATATCTGTATACCACCACTCCCTAGGG;CAGGAG 17338		
Oy	674 gcacaataagaaagacagcatcacygccatgctgtaagctcygcagaagg;tctctt 733		
Db	17337 GCTGCATTAGTGAACATGTGTCTGACACAGCATCAGCATCAGATGTAACACACCAAGTGCTGT 17278		
Oy	734 tctctctgatgctgcacaatgatatccaagccccagaacatcgcagaatctcgaag :octga 793		
Db	17277 TCCTGTGGATGATCAACAAGTGGGCT---GCCCTCACTCCCCAAGCCCTACACAA;ACTGA 17221		
Oy	794 taaagaaaaaacacgcgtctcaagaacatbgutcatctgcaccactlcaaatgtagt:cctga 853		
Db	17220 TTACAAAATACTACTTGTTCGCGACCTGCTTATGTGCTCTGTTCATACAAACAG;GTGA 17161		
Oy	854 ggacatacggc 865		
Db	17160 GAGACATCCGCC 17149		
RESULT 15			
AF135490	4752 bp mRNA	ROD	29-Sep-1999
LOCUS	Mus musculus clone Z1 neuronal apoptosis inhibitory protein (Naip)		
DEFINITION	mRNA, complete cds.		
ACCESSION	AF135490		

[illegible]

Db 1707 CATCAGATGCTGCCCCCTGTGAAACAGGTTCCACGTGCTTCTTACCTTCCTTAGTT 1766
Oy 614 tcaagcaggccccagggtgacttttgaaccctctgtgatacactcctlgatataccctg 673
Db 1767 CCATCACACCAAGCCAGGCACTGGCCAAATATCTGTGCCCACTCCTAGGGGCGAG 1826
Oy 674 gcacaatcaggaaagcagacatcaltgccaatgctgtaagcctgcgcagaaaggttctt 733
Db 1827 GCTGCATTAGTGAAGTGTGTCTGAGCAGCATCATCCAGCAGTTACACACCAAGTGTGT 1886
Oy 734 tccttcttgatggctacaatg 754
Db 1887 TCCTGTGATGACTACAGTG 1907

Search completed: March 25, 2002, 13:24:16
Time: 9240 sec

THIS PAGE BLANK (USPTO)


```

QY 323 LQIQSRCLNLMKTPLEFVITCALOMGESEFHSHTOTLFTFTYDILLQKN HKHKGVA 382
Db 627 VYFGKQSLQKLOKTPLEFVAALCAHMFQYPRDPSPDVAVFYSMERLSLRN-----A 680
QY 383 ASDFLR-SLDHGGDLALLEGVFSHKDFELOY--SSVNDVLLTGLCKYTA JRFKPKY 439
Db 681 TAEILKATVSSCGELALKFSCCFEFDNDLAEAGVDEDEDLTKLMSKFTYJRLRPFY 740
QY 440 KFFHKSFOETAGRLSSLTSHPEPEVTKGNGYLQKMWISIDITSTYSLLR ITGGSV 499
Db 741 RFLSPAFQEFGLMRLIELDSDROHDGLYLHLKQINSPMTAVANNFLTV--SSL 798
QY 500 EATRA---VMKHLAAVYOHGCLGLSIAKRLPMLROESLOSVDN-----TTEOET 545
Db 799 PSTKAGPKIVSHLLHLYDN-----KESLEISENDDYLKHOPE SLQMOVL 843
QY 546 LKAI-----NINSFVECCGILH-----YQESTSKSALSOEFAPFGQSLYIT SGNIPD 593
Db 844 LRGLMQLCPQAFVSNVSEHLVYALKTAVQSN--VAACSPFVLQFLQGRLLTGLALNTL-Q 901
QY 594 YLPD-----FPEHLPLNC--ASALDFIKLDFYGG--AMASWE 625
Db 902 YFPDHPESLSLRSHFSIRGNKTSPPRAHFSVLETCFQKSVPTIDQVYASAFEPNENE 961
QY 626 -----KAEPTGGIHMEABEY--IPSRVSLFFNNKQEFRTLEYVL 666
Db 962 RNLAEKEDWKSYMOMGRASPDLSTGYWKLSPKQYKIP-----JLEVDY 1006
QY 667 RQFSKLKODITYLKGIFSSATSLRLQIKRCAGVAGSLVSTCK-NIYSLWASPLT 725
Db 1007 NQDIDVGGQMLELMTVFSASQRIELHNSRGFTESIRPALELSKASVTKSISIKLELS 1066
QY 726 IEDERHITSYTNKLTLSIH---DLQNRPLPGCLTDSLGNLKNLTKLIMDNIRK-- 775
Db 1067 AABQELLTLPLSLEVSQTIOSODQIFPN--LDKFLCKELSLVDLEGINVFVIPEE 1124
QY 776 -----NEEDAIKLAEGLKLNKKKCLFHLTHL---SDIGEMDYIVKSL 816
Db 1125 FPNFHHMEKLLIQISAEYPSKLVASLPNLSIKLILNEGOQFPDEETSEKAVYJLGSLS 1184
QY 817 SEPCDLEELQVSCCLSANAVKILAOHLNLVYKLSIL 853
Db 1185 ---NLEELLPTGDIYVAKLILIQCCOCLHCLRLV 1217

RESULT 2
US-08-836-134-23
Sequence 23, Application US/08836134A
Patent No. 6020127
GENERAL INFORMATION:
APPLICANT: Mackenzie, Alex E.
APPLICANT: Korneluk, Robert G.
APPLICANT: Mahadevan, Mani S.
APPLICANT: McLean, Michael
APPLICANT: Roy, Natalie
APPLICANT: Ireda, John
TITLE OF INVENTION: Neuronal Apoptosis Inhibitor Protein, Gene Sequence and
Patent No. 6020127
TITLE OF INVENTION: Mutations Causative of Spinal Muscular Atrophy
FILE REFERENCE: 3477-112, 033477/139914
CURRENT APPLICATION NUMBER: US/08/836,134A
CURRENT FILING DATE: 1997-06-20
NUMBER OF SEQ ID NOS: 23
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 23
LENGTH: 1151
TYPE: PRT
ORGANISM: Homo sapiens
US-08-836-134-23

```

Query Match 10.2%; Score 540.5; DB 3; Length 1151;
Best Local Similarity 24.5%; Pred. No. 3.1e-40;

```

Matches 204; Conservative 144; Mismatches 300; Indels 185; Gaps 30;
QY 69 FLKSLKEMNYPLFQDINGO-----SLFQTSSEGLDD----- 100
Db 353 FQNKMS--SAEYTPDLOSNGEJLCELETTSSESNLEIAGVPIVPEMAQGEAFQEOAKN 411
QY 101 LAQDLKDLHYTBSFLNF-----YPLGEDIDITFNKSTFTEPVLKRKQDHNHR 148
Db 412 LNEQLAAATVTSASFRRMSLLDISDLPATDHLGCDIST--ASKHISKV----- 458
QY 149 VQDLTLNGLLQALQSPCLITEGESGKRSLLQRIAMLMGSGCKALTKRFVFLRLS-- 206
Db 459 QEFLVLPVEFGNLSVWCVEGAGSGKTVLLKRIAFMLMSGCCPLLNRFQVFLYLSST 518
QY 207 RAQGLFETLCQDLDDPTIKQTFEMAMLLKRLQRYFLDLGYNF--KPNQCEIEAL 264
Db 519 RPDEGLASITICQDLLEKESVTEMCNRNIIQQLKNQVFLLDYKREISIP--VIGKL 575
QY 265 IKENHFRKMVAVTTTECLRHIFRGALTLAEVGMTEDSAQALIREVLIKELA--EGLL 322
Db 576 LQKNSLRCLLAVNTNRADIRRYLETILQAFEPFYNYCIRKLFSSHMTGLRKM 635
QY 323 LQIQSRCLNLMKTPLEFVITCALOMGESEFHSHTOTLFTFTYDILLQKNHKHKGVA 382
Db 636 VYFGKQSLQKLOKTPLEFVAALCAHMFQYPRDPSPDVAVFYSMERLSLRN-----A 689
QY 383 ASDFLR-SLDHGGDLALLEGVFSHKDFELOY--SSVNDVLLTGLCKYTAQRFKPKY 439
Db 690 TAEILKATVSSCGELALKFSCCFEFDNDLAEAGVDEDEDLTKLMSKFTYJRLRPFY 749
QY 440 KFFHKSFOETAGRLSSLTSHPEPEVTKGNGYLQKMWISIDITSTYSLLRITGGSV 499
Db 750 RFLSPAFQEFGLMRLIELDSDROHDGLYLHLKQINSPMTAVANNFLTV--SSL 807
QY 500 EATRA---VMKHLAAVYOHGCLGLSIAKRLPMLROESLOSVDN-----TTEOET 545
Db 808 PSTKAGPKIVSHLLHLYDN-----KESLEISENDDYLKHOPEISLQMOVL 852
QY 546 LKAI-----NINSFVECCGILH-----YQESTSKSALSOEFAPFGQSLYINSNIPD 593
Db 853 LRGLMQLCPQAFVSNVSEHLVYALKTAVQSN--VAACSPFVLQFLQGRLLTGLALNTL-Q 910
QY 594 YLPD-----FPEHLPLNC--ASALDFIKLDFYGG--AMASWE 625
Db 911 YFPDHPESLSLRSHFSIRGNKTSPPRAHFSVLETCFQKSVPTIDQVYASAFEPNENE 970
QY 626 -----KAEPTGGIHMEABEY--IPSRVSLFFNNKQEFRTLEYVL 666
Db 971 RNLAEKEDWKSYMOMGRASPDLSTGYWKLSPKQYKIP-----CLEVDY 1015
QY 667 RQFSKLKODITYLKGIFSSATSLRLQIKRCAGVAGSLVSTCK-NIYSLWASPLT 725
Db 1016 NQDIDVGGQMLELMTVFSASQRIELHNSRGFTESIRPALELSKASVTKSISIKLELS 1075
QY 726 IEDERHITSYTNKLTLSIH---DLQNRPLPGCLTDSLGNLKNLTKLIMDNIRK 775
Db 1076 AABQELLTLPLSLEVSQTIOSODQIFPN--LDKFLCKELSLVDLEGINV 1126

RESULT 3
US-09-180-439-3
Sequence 3, Application US/09180439
Patent No. 6225532
GENERAL INFORMATION:
APPLICANT: Dixon, Mark S
APPLICANT: Hatziathanis, Kostas
APPLICANT: Jones, David A
APPLICANT: Jones, Jonathan DG
TITLE OF INVENTION: Plant pathogen resistance genes and uses thereof
FILE REFERENCE: 620 - 53
CURRENT APPLICATION NUMBER: US/09/180,439
CURRENT FILING DATE: 1998-12-06
EARLIER APPLICATION NUMBER: PCT/GB97/01249

```

[illegible]

RESULT 4
00-180-439-4
US/09180439

US-09 Sequence 4, Application: 62255310
 Patent No. 62255310
 GENERAL INFORMATION:
 INVENTOR: Mark S
 APPLICANT: Dimitriou, Kostas
 APPLICANT: Hatzixanthidis, Kostas
 APPLICANT: Jones, David A
 APPLICANT: Jones, Jonathan DG
 APPLICANT: Jones, Jonathan DG
 TITLE OF INVENTION: Plant pathogen resistance genes and uses thereof

[illegible]

RESULT 5
08-030-996A-2
US/08930996A

US-08-930-27
Sequence 2, Application
Patent No. 6100449
GENERAL INFORMATION:
APPLICANT: FUHR, Robert

APPLICANT: ESHED, Yuval
APPLICANT: ORI, Naomi
APPLICANT: PAPAN, Ilan
APPLICANT: ZAHAR, Daniel
TITLE OF INVENTION: A GENE FAMILY FROM THE 12 FUSARIUM RESISTANCE
TITLE OF INVENTION: LOCUS OF TOMATO AND USE THEREOF FOR TRANSFORMATION AND
TITLE OF INVENTION: SELECTIVE BREEDING OF TOMATO, AND RELATED PLANTS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEWMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/930,996A
FILING DATE: 09-DEC-1997
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/05272
FILING DATE: 15-APR-1996
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: IL 113,373
FILING DATE: 13-APR-1995
INFORMATION FOR SEQ. ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1220 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-930-996A-2

Query Match 32% Score 170.5; DB 3; Length 1220;
Best Local Similarity 19.8%; Pred No 2.4e-06;
Matches 231; Conservative 160; Mismatches 393; Indels 383; Gaps 61;

DB 9 RALIQKMGIVYQITDDLVNANVLRNEEYNIICCEVEDDARAGIIMLLKSGESCN- 67
242 KGLIQEIGSTDLK-ADD-----NLNOLVKKLADNMLNO-----LQV-LKEKLG 285
68 -LEKSLKE-W-NYPLFODLNGSLFHOTSEGLDD-----LAQ-LKDLVHT 111
286 KRLPLVLDVWMDNVPWMDL--RNLFLO--GDIGSKITVTRKESVALMDSGALY-- 338
112 PSLNLFYPLG---EDIDILFNKSTFTEPVLMRKD-QHHHRYEQL-----T-NGI--- 157
339 -----MILSSEDEWALFKRHS-----LEHKDKEPEDEEVGKQIADK KGLPLA 384
158 LQALQSPCIIIEGSGKSTLQRIAMLWGSCKALTKRFVFFLRLSBAQGGLEFLLC 217
385 LKALAGLRKSEVDEWNRILRSEI--WELPSC-----SNG-LPALM 424
218 DQLDLIPGTRKQTFMAMLLKRLQRYLFLLDGYNEFRPQNCPELEALIKENHR- KMMVT 276
425 LSYNDLPRANKO-----CLAVCAIYPRDYQFRKDOVI 456
277 VTTTTECLRIIROFGA-----LTAEGVDMTEDSAQALIREVLIKELAGLLAIOKS 328
457 HLMIANGLVH--OFHSGNOYFIELRSRSLFEMASEPSEBVEEFLMDLVNDLAQIASSN 514
329 RCLNKLKTPLEFVITCAI-----QMGESB-----FHSHTQTL-----FHTFYLLI 371
515 HCLRLDENKSGHMLQCRHMSYSTIGDGEFEKILSLFKSGLRLTLPLIDIOFH-YSLKLS 573
372 OKNKHKRGVAASFIRSLDHC-----GDALLEGVSHKFDLQOVSSVNEQVL----- 421
574 KRVYHNHLPRLSRALSLSHYQIEVLPNDLFIKILKLRFLDLSETSIITKLPSIFV.XN 633

DB 422 LTTGLL--CKYTAQKFKPKYKFFHKSPQETTAGRLS--SILTSHEPEVYKNGYL--- 474
634 LETLLSCEVLEEDPLQMEKILINHLIDISNTRKRLPMLHSRLKSLQVLGAKFLVG 693
475 OKRVASIDTTSYSSLLRYTCSSV---EATPAVKHILAAYQOCLLOSLAKPRLRQ 531
694 WRREYLGEPNLYGSLSTILELNVDRRAVAKAKREKNHVO-----LSLE---W-S 742
532 ESLOSVMKNTTEQEL-----KAININSVEGCIHLYOESTSKSALSQEFEEFQOG 582
743 ESTSADNSQTERDILDELPRKNIRAVEITGY----- 774
583 SLYNSGNIPLYLFD-----FEHLPNCAS-----ALDFIKL----- 614
775 ---RCITNPNWVADPLFVKILVHLRLNCKDCYSLPALGQLPCLFELSRGMGIRVYTE 830
615 DEYCGAMASMEKAEDTGIHMEAPETYIPSAVSLFFNMKQ-----EERTLEVTLR 667
831 ERYG--RLSSKRPNSLVYLRFEEDME-----WKQWHTLIGIEPPTLE--- 871
668 DFKLKNQODITLGIKIFSSATSLRLQIKRACVAGSLSVLST-----CKNIYSIMV 719
872 -----KLSIKNCPELSLEIPGFSSLRDLIDCDKSVTSFPF 908
720 EASPLTIE-----DERHITSVNTKTLISHLOLQORLPGLGILDSLGMLK 763
909 SLPTTLKRIKISGCPKLEAPVGMFEVYLSVLDCCGVDDISPEFLPTARQLSIENCH 968
764 NTKLIMDNIKNMEEDAIKLAGIKRLKMKCL-----PHLTHLSDIGEMDIYVKSLSSE 818
969 NVRFLPIPTATESLH-----IRNCEKLSMACGGAQDTSLINWG-----CKKLCL 1014
819 P--COLEFIOIYSCGLSANAVKILIAONLHNLYKISTIDLSSENYLER--DGNEALHELID 873
1015 PELLPLKELRLTYC-----PEIBELPFN--LOIDI--RYCKKLNGKREWH--- 1059
874 RMNVLBOLTAIMP--GCDVQGSLSLKHLE--EYPOLYKLGKLNWRLTDEIRLAF 930
1060 ---LDRLEMLKTHGSD-----EHLEWELPSSIQ-----RLFTINIKLSQ 1100
931 FGNPNLKNQOQLAQN--RVSSDGLWLMGVENIKOLVDFDFSKRELPLPALYKRS 988
1101 HLKS-LTSLQPLRTVGNLSQFQSGQLSFSHLTSLQTLQIMWFLNQLSPSALPSSLS 1159
989 Q-VLSKLTFLQEARLVGQFDDDLASV 1014
DB 1160 HLISNCPNLSLPLKGMPSLSLSTLSI 1186

RESULT 6
US-09-353-585-2
Sequence 2, Application US/09353585
Patent No. 6287865
GENERAL INFORMATION:
APPLICANT: Dixon, Mark S
Jones, Jonathan DG
TITLE OF INVENTION: Plant pathogen resistance genes and uses
thereof
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon & Vanderhye PC
STREET: 8th Floor, 1100 No. 6287865th Glebe Road
CITY: Arlington
STATE: Virginia
COUNTRY: United States of America
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/353,585
FILING DATE: 15-Jul-1999
CLASSIFICATION: C12N 15/29, 15/82, A01H 5/00, A01N 65/00, C12O 1/68
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/930,277
FILING DATE: 27-OCT-1997
APPLICATION NUMBER: PCT/GB96/00785
FILING DATE: 01-APR-1996
APPLICATION NUMBER: GB 9506658.5
FILING DATE: 31-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Ms Mary J Wilson
REGISTRATION NUMBER: 32,955
REFERENCE/DOCKET NUMBER: 620-69
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1112 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Tomato
STRAIN: Cf2
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-353-585-2

Query Match 3.1%; Score 165; DB 4; Length 1112;
Best Local Similarity 21.6%; Pred. No. 6.6e-06;
Matches 139; Conservative 81; Mismatches 199; Indels 224; Gaps 31;
425 GLICKYTAQRKPKKFFHRSFOETAGRRRLSLTSHPEEYTKGNGYLOKMWISDIT 484
140 GLALAQIR-----IFHNO-----LNGFPEI-----GYLSRLKLS-LG 175
485 STYSILRYTCGSVEATRAVMKHLAAVYQHCGLGISIAKRPLMROESLOSXKTTBOE 544
176 INFLS-----GSIPASVGNLNNISFLYLYNNQSSSI-----PEISYLSNLTPLD 221
545 ILKAININSVECGIHLQESTSKALSQEFDA-----FPGKSLYINSGNIPD 593
222 L-----SDNALNGSIPASISGNNNNLSFLYGNQL---SGSIPE 257
594 ---YL-----FDPEHLPN---CASALDFIKLDF---YGGAMASWEKAEDTGGIHMEEA 639
258 EICYURSLTYIDLSENALNGSIPASISGNNNNLSFLYGNQLSG--SIPEIG----- 308
640 PEYIPSRVSLFPMKOEFTLEVTLLRDFSKLNKODITYLGKIFSSATSLRLQIKRCAG 699
309 ---YL-----RSLNVLGISENALN----- 324
700 VAGSISLVLSTCKNITYSLMEASPLTJEDERHTSVNLTSLHDLONQRLPGGLTDSL 759
325 --GSIPASISGNNISRLNLTNNQSSGSIPEIGSLTYIDL 379
760 GNLNLTFLINDNFKMNEDEAKLAEGKLNKKMCLPHLTHLSDIGCEMDIVT-----SL 815
380 GNLNLSMLYLYNNQSSGSIPEIGSLTYIDL 438
816 SSEPCD-----LEEIQVSCCLSANAVKILAAONLHNLVLTSLIDSEN 858
439 SNNISNGIFPASFGNMSLAFLEYENQLASS-----VPEIGYLRSLNVLDSLEN 489
859 YLEADGNALHELDKMANVLEQLTALMLPWCCDVQGSISLILKHEEYPPQLYKGLKNWR 918
490 AL---NGSIPASISGNNISRLNLTNNQSSGSIPEIGYLRSLNVLDSLEN 533

919 LTDREIR-ILGAFPGKNPLKFNQOLNLAGNRVSD-----CWL----- 955
534 LSENALNGSIPASFGN--LNNLSRLNLTNNQSSGSIPEIGYLRSLNVLDSLENALNGSI 591
956 -AFMGVPENLKQLVFDFSTKEFLPDPAIVRKLSQVLSKLTFL 997
592 PASISGNNISRLNLTNNQSSGSIPEI-----GYLSLTYL 628

RESULT 7
US-09-353-585-3
Sequence 3, Application US/09353585
Patent No. 6287865
GENERAL INFORMATION:
APPLICANT: Dixon, Mark S
Jones, David A
TITLE OF INVENTION: Plant pathogen resistance genes and uses thereof
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Nixon & Vanderhye PC
STREET: 8th floor, 1100 No. 6287865th Glebe Road
CITY: Arlington
STATE: Virginia
COUNTRY: United States of America
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/353,585
FILING DATE: 15-Jul-1999
CLASSIFICATION: C12N 15/29, 15/82, A01H 5/00, A01N 65/00, C12O 1/68
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/930,277
FILING DATE: 27-OCT-1997
APPLICATION NUMBER: PCT/GB96/00785
FILING DATE: 01-APR-1996
APPLICATION NUMBER: GB 9506658.5
FILING DATE: 31-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Ms Mary J Wilson
REGISTRATION NUMBER: 32,955
REFERENCE/DOCKET NUMBER: 620-69
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1112 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Tomato
STRAIN: Cf2
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-353-585-3

Query Match 3.1%; Score 165; DB 4; Length 1112;
Best Local Similarity 21.6%; Pred. No. 6.6e-06;
Matches 139; Conservative 81; Mismatches 199; Indels 224; Gaps 31;
425 GLICKYTAQRKPKKFFHRSFOETAGRRRLSLTSHPEEYTKGNGYLOKMWISDIT 484

Tue Mar 26 11:36:23 2002

us-09-697-089-2.ra1

Page 6

Db 140 GILAKLOIR-----IFHNO-----LNGFIPKEI-----GYLRCTKLS-LG 175
OY 485 STYSSLLRYTCGSSVEATRAVMKHLAAVYOHGCLGLSTAKRPMQESLOS/KMTTQOE 544
Db 176 INFIS-----GSTPASVGNLNNLSFLYLYNNQSGSI-----PEEIST--NSLTEL 221
OY 545 ILKAININSVEGCIHLVQESTSKALSQDFEA-----PFOCKSLX--MSGNIP 593
Db 222 L-----SDNALNGSIPASIGNMMNLSFLYLYNNQSGSI-----SGSIPE 257
OY 594 ---YL-----FDFEHLPN---CASALDFIKLDF---YCGAMASWEKAEPTGCIHMEEA 639
Db 258 EICYLRSLAYLDSEVALNGSIPASIGNMMNLSFLYLYNNQSGSI-----SIPEIG 308
OY 640 PEYIYPSRAVSLPFNNKQEFRTLEYTLRDFSKLNKODITYLGKIFSSATSILRIJIRKAG 699
Db 309 ---YL-----KSLNVLGISENALN-----
OY 700 VAGSLSLVSTCKNTKNTYSIMVEASPLTIEDERHITSVNLKTLSTIHDONORLP:GLTDSL 759
325 --GSIPASIGNKLNNLSRLNLYNNQSGSI-----ASIGLNNLSMLYLYNNQSGSI:SI 379
OY 760 GNKMLTFLINDNIRKMEEDAIRKAEGLKMLKMCIFHLTHLSDIOGMDYIV:---SL 815
Db 380 GNLNLSMLYLYNNQSGSI:SI-----NLNNLSMLYLYNNQSGSI:PEEIGYLS:LYIDL 438
OY 816 SSEPQD-----LEFIOLVSCCLSANAVKIIAQNLMNLVYKSLIDISEN 858
Db 439 SNNISNIFIPASFGNMSNLAFLFLYLYNNQSGSI-----VPEEIGYLSRLNLYNNQSGSI 489
OY 859 YLEKDGHEALHELIDRMNVLEQTLALMLPWGCDVQSGSILKHEEVPOLVKT:ILKNWR 918
Db 490 AL-----NGSIPASFGNMMNLSRLN-----VNNQSGSI:PEEIGYLS:LYIDL 953
OY 919 LDTETIR-ILGAFEGKNPLKNFOQLNAGRYSSD-----GWL----- 955
Db 534 LSNALNGSIPASFGN--LNNLSRLNLYNNQSGSI:PEEIGYLS:LYIDL 955
OY 956 -AFMGVFNKQVLEFDESKTEFLPPALVLRKLSOVLSTLTEL 997
Db 592 PASIGLNNLSMLYLYNNQSGSI:PEEIGYLS:LYIDL 628

RESULT 8
US-09-180-439-8
Sequence 8, Application US/09180439
Patent No. 6225532
GENERAL INFORMATION:
APPLICANT: Dixon, Mark S
APPLICANT: Hatzixanthis, Kostas
APPLICANT: Jones, Jonathan DG
TITLE OF INVENTION: Plant pathogen resistance genes and uses thereof
FILE REFERENCE: 620 - 53
CURRENT APPLICATION NUMBER: US/09/180,439
EARLIER FILING DATE: 1998-12-06
EARLIER APPLICATION NUMBER: PCT/CB97/01249
EARLIER FILING DATE: 1997-05-08
EARLIER APPLICATION NUMBER: GB 9609681.3
EARLIER FILING DATE: 1996-05-09
EARLIER APPLICATION NUMBER: GB 9619924.5
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 8
LENGTH: 1016
TYPE: PRT
ORGANISM: Lycopersicon esculentum
US-09-180-439-8

Query Match 3.1%; Score 164; DB 4; Length 1016;
Best Local Similarity 22.1%; Pred. No. 7e-06;

Matches 147; Conservative 115; Mismatches 222; Indels 182; Gaps 38;
OY 459 LPSHEEYKNGYLGQWVSI--DITSTYSSLRITCGSSVATRAVMKHLAAVYOHG 517
Db 203 LSGFIPEI-----GYLRSLTKLSIDIN-----FLSGISIPASIGNMMNLSFLYLYNN 249
OY 518 ILGLSIAR-----PPLWR-----QESIOSVKNTRDEI----- 545
Db 250 QLSGSIPEEIGYLSRLKLSIGLNFISGSI:SI-----LNNLSLIDLNKLSI:PEEIGY 309
OY 546 LKAININSFVDCGCIHLVQES--TSKALSQDFEAFPOGSLYNSGNIPD---YL----- 595
Db 310 LRSI-----TYLDGEMALNGSIPASIGNMMNLSFLYLYNNQSGSI-----SIPEIGYLS 639
OY 596 FDFEHLPN-----CASALDFIKLDFYCGAMASWEKAEPTGGIHM-----EAPET 642
Db 364 LDGEMALNGSIPASIGNMMNLSRLDLYNNKLSG--SIPEEIGYLSLYLDGEMALNG 421
OY 643 YIPRAVSLPFNNKQEFRTLEYTLRDFSKLNKODITYLGKIFSSATSILRIJIRKAGVAG 702
Db 422 SIPASIGN-----NNLPMYLYNNQSGSI:PEEIGYLS:LYIDL 469
OY 703 SLVSTCKNTKNTYSIMVEASPLTIEDERHITSVNLKTLSTIHDONORLP:GLTDSL 759
Db 470 SIPASIGNMMNLSFLYLYNNQSGSI:PEEIGYLS:LYIDL 523
OY 760 GNKMLTFLINDNIRKMEEDAIRKAEGLKMLKMCIFHLTHLSD--TGEOMDYIVSL 815
Db 524 GNLNLSMLYLYNNQSGSI:SI-----GSIPASFGNMMNLSRLN-----FLSDNDLGEIPSVCNLT 858
OY 816 SSEPQD-----PC--DDEIOLVSCCLSANAVK--IIAQNLMNLVYKSLIDISEN 907
Db 576 SLEVLVYNNRNLKQKVPQCGNISDLHIS--WSSNSFEGELPSSISUTSIKILIDFGRN 907
OY 859 YLEKDGHEALHELIDRMNVLEQTLALMLPWGCDVQSGSILKHEEVPOLVKT:ILKNWR 918
Db 634 NLEBAPFOFGNIS:SIQVDFEOMN--NKLSG--TLPTNFSIGSLSLIHLNGELADEIPRS 950
OY 908 ---OLVRLGILKNNRLTDT-----EIRIIGAFGR--NPLK-----NEQOL 942
Db 691 LDNCKRLQVLDGDNLDNDTFFPMWGLTLPBLRVNLTSNKLHGFRSSGAIMPPDLRII 750
OY 943 NLAGNRYSSDGMIAFMGVFNKQVLEFDESKTE-----FLPPALY-----RKLSOVL 991
Db 751 DLSRNPQD---LPTSLFEHLKGMRTVDKTMEPPESEYVDSDSVYVYKGLFELTVRL 807
OY 992 SKTEFL 997
Db 808 SLVTVI 813

RESULT 9
US-08-353-700-1
Sequence 1, Application US/08353700
Patent No. 5599919
GENERAL INFORMATION:
APPLICANT: YEN, TIMOTHY J
APPLICANT: RATTNER, JEROME B
TITLE OF INVENTION: NUCLEOTIC ACID ENCODING A
TITLE OF INVENTION: TRANSIENTLY EXPRESSED KINETOCHORE PROTEIN,
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: DANN, DOREMAN, HERRELL AND SKILLMAN
STREET: 1601 MARKET STREET, SUITE 720
CITY: PHILADELPHIA
STATE: PA
COUNTRY: USA
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS


```

* SOFTWARE: PatentIn Release #1.0, Version #1.25
* CURRENT APPLICATION DATA:
* APPLICATION NUMBER: US/08/353,700
* FILING DATE: 09-DEC-1994
* CLASSIFICATION: 435
* ATTORNEY/AGENT INFORMATION:
* NAME: REED, JANET E.
* REGISTRATION NUMBER: 36,252
* TELECOMMUNICATION INFORMATION:
* TELEPHONE: (215) 563-4100
* TELEFAX: (215) 563-4044
* INFORMATION FOR SEQ ID NO: 1:
* SEQUENCE CHARACTERISTICS:
* LENGTH: 3248 amino acids
* TYPE: amino acid
* STRANDEDNESS: single
* TOPOLOGY: linear
* MOLECULE TYPE: protein
* HYPOTHEetical: NO
* ANTI-SENSE: NO
* ORIGINAL SOURCE:
* ORGANISM: HUMAN
US-08-353-700-1

```

Query Match 2.8%; Score 149.5; DB 1; Length 3248;

Best Local Similarity 18.4%; Pred. No. 0.001; Matches 213; Conservative 195; Mismatches 461; Indels 287; Gaps 52;

```

QY 39 NICECEVEDADAGIIMILKSGESC-----NLFKSLKENWYPLFODLNGO-- 87
D 684 NVLDSKSVETQK-LAYMELQOKAEFSDQKHQKEIEMCLKT-----SULTGVE 733
QY 88 SLFH--QTSBGDDDLADLDLYHTPSLFNYPGEGIDIIFN-----LKSFTPEVL 139
D 734 DLEHLQLLSNEIMDKRCYQDLHAYESRLDLKSKDASLVTDHORSGLAFDQOPAM 793
QY 140 WRKDOHHRVBQTLNGLQALQSPCIEGSGKSTLLQ-RIAMLGSGCKALTRFK 198
D 794 -----HSEFANITIGEGSMPESEKRELDQSPKNSAILQNRVDSL----- 835
QY 199 FVFLRLSRAGGLFETLDCQDLDPGTIRKQTFMAMLLKLRQVFLFDLYNEFKPQNC 258
D 836 -EFSLESGKQNSDLQKCELVQIKGEIEN-----LMKAQK-----HQSFVETS 882
QY 259 PEIALIKENHRFNMYIVTTTCELRHIFQFALTAEGVDMTDSQAOLIREV-LIKEL 317
D 883 QRIKLEQEDTSAHQ-VVAETLSALENKEKELQLLNDKV-----ETQAEIQELKKS NHL 936
QY 318 AEGLLQIQKRCRLNMLKPTLFAVITCAIQMGSEFHSHTQ--TLFHTFYDL----- 369
D 937 LEDSLKELQLLSETLSLEKKMSSII-----LNRKEI BELTQENGTLKETINASINQEKMN 992
QY 370 LIQRNK-----HKHGVN-ASDFIRS-----LDHGGDL-ALEGVSHKFEDELQVS 414
D 993 LIQKSESPANYIDREKSISLSQYQKQKILLQRCETGNAYEDL-SQYK-AQEK 1050
QY 415 SVNEDVLTTLCKYTAQRKPKYKFFHKSFOY-----TAGRRL 455
D 1051 SKLECLLNCTSLCENRKNLEQLKEAFKHEQFLTKLAFERRNONLMLELTVOAL 1110
QY 456 SSLTSHPEPVTGNGVLOKQVSI-----ITSVSLRLRYTGSSVEATRAVMKHLAA 511
D 1111 RSEMTDONNNSKSPAGGLKQIMTLKEQNKQKQEVNDLQEN-----EQLMKVK----- 1161
QY 512 VYHGCLGLSIAKPLMROBSIOS--VKNTTEQELIKAININSFVCGIHLVQES-TSK 568
D 1162 -TKHEC-----ONLESEPIRNSVKRESEBNQCNFKPQMDLEVKELISDSY 1206
QY 569 SALQGFPAFFQSGSLYINSINIDYLFDPFEHLPNCAS-ALDIKIDPFGGAMASKEKA 627
D 1207 NAOVLQEAAMLKRELQES-----EKKECLQELQETIRQDLTSLNDQMO-- 1254

```

```

QY 628 AEDTGIMMER--APETYIPS-RAVSLFFNMKQEF--TLEVTLRDFSKLNKQDITLGI 683
D 1255 SQEISGLKDCIIDEKRTISGPHSLSTQNDNALQSLQTTMKLNELEKICETLOAK 1314
QY 684 FSSATSLRLQIRKAGVAGSLVSTCKNITYSLMVASPL-----TIED----- 728
D 1315 YELVTELNDRSSECTITATRKMAEEVQKLLNEVKILNDQSGLLHGEIVEDIPGGEFGQPN 1374
QY 729 ERHITSVTNL-----KTLSHDIQONRLPGGLTDSIGNIKNLTKLMD----- 771
D 1375 BQHPVSLAPIDESNSYEHITLTDKEVQMHFAELOEKFLTSQSEHKILHDQCMSSKMS 1434
QY 772 -----NIKMERDAIK-----LAEGI-KNLKMCFLHLTHSDIEGM 808
D 1435 LQTVYDSLKAENLVSLTNLRFQGLVKEQMLGEBGLVPSLSSCVDPSSSLSSLDSS 1494
QY 809 DY-----IVKSLSEPCDLEIEIQLVSCCLSNAYK-----ILAQNLHN--- 846
D 1495 FYRALLDQTDGMSLSLNIEGAVSANQSVDEVPSSLOTVYDSLKAENLVSLTNLRFQ 1554
QY 847 -LVKLSLIDSENVL-----EKDQNEALHELIDRMNVLEQTLALMP 887
D 1555 DLVKEMQGLEEGLVPSLSSCVDPSSSLSSLDGSSFYRALLDQTDGMSLSLNIEGVSA 1614
QY 888 WGDVQGSLSLSLKH-----LEEVPQLVKLGKNNRLDTFELRLGAFGKNPLKNPQOLN 943
D 1615 NQGVDEVPSSLOEENULTRKETSPAPAKVE-----ELESICEYR-----OSLE 1660
QY 944 LAGNRVSSDGLAFMGVFNILKQLVF-----FDSKEFLPDPALVYKLSQVLSKTLFIQ 998
D 1661 KLEKMSQGIKMKKEI-QELEQQLSSEROELDCLRRQYISEN---EQWQKLTLSVILEM 1716
QY 999 EARLVNQFDDDLSTV 1014
D 1717 ESKLAERKQTEQLSL 1732

```

RESULT 10

Sequence 1, Application PC/TUS9516216

GENERAL INFORMATION:

APPLICANT: Yen, Timothy J.

APPLICANT: Ratner, Jerome B.

TITLE OF INVENTION: Nucleic Acid Encoding a Transiently

TITLE OF INVENTION: Expressed Kineochore Protein, and Methods of Use

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Dann, Dorfman, Herrell and Skillman

STREET: 1601 Market Street Suite 720

CITY: Philadelphia

STATE: PA

COUNTRY: USA

ZIP: 19103-2307

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/16216

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/353,700

FILING DATE: 09-DEC-1995

ATTORNEY/AGENT INFORMATION:

NAME: Reed, Janet E.

REGISTRATION NUMBER: 36,252

TELECOMMUNICATION INFORMATION:

TELEPHONE: (215) 563-4100

TELEFAX: (215) 563-4044

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 3248 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: Protein
HYPOTHEITICAL: NO
ANTI-SENSE: NO
PCT-US95-16216-1

Query Match 2.8%; Score 149.5; DB 5; Length 3248;
Best Local Similarity 18.4%; Pred. No. 0.001;
Matches 213; Conservative 195; Mismatches 461; Indel: 287; Gaps 52;

39 NICEKVDADARGIHHILKSGS-----NLFKSLKEMNYPIFODINGQ-- 87
684 NYLDSKSYVEVETOK-LAYMELQOKAFESDQKHQKEIENMCLKT-----SOLTGQVE 733
88 SLFH--QTSEGDLDLAODKLDLYHTSPFNFPYPLGEDIDIDFN-----LSTFTEPVL 139
734 DLEHKIQLLSNEIMDKRCYQDLHAEYSRLDLKSKDASLVTNEDHQSL.AFDQOPAM 793
140 WRKDOHHNVEODLTNGLLQALOSPCTIGESGKSTLLO-RIMALWMSGI CKALTKFK 198
794 -----HSPANIIIGEGSMSPSERSECRLEADQSPKNSAILQNNVDSL----- 835
199 FVFFELSLSAOGSLPFTLDDLDIPGTIRKOTFMAMILKLRVLFLLDGYKEEPKQNC 258
836 -BESLESOKOMSDLOKCEELVOIGELIEN-----LKKACDM-----HYSFAETS 882
259 PEFEALIKENHKFNMYVITTTTECLRHTRQGCALTAEGDMTEDSAQALIRV-LIKEL 317
883 ORISKOEDTSAHQ-VAVETLSALENKEKELQDLNDKY-----EEOQAEIQI LKKSNNH 936
318 AEGULQIQOKSRCLRLMTPLFVYVTCALOMGESFHSHTOT-TLPHFTYFL----- 369
937 LBSLSKELQDLSTLSEKEMSSITIS--LKKREIELIQLQNGKLKELINSUNOEKKN 992
370 LIOKKN-----HKHGYA-ASDFTRS-----LDHCGDL--ALBESVSHKED TELQDVS 414
993 LIOKSESFANYDERKRSISELSDQYKOEKLIILORCEEGNAVEPL-SQKTK-AAQOEN 1050
415 SVNEDEVLLTGLCKYTAQRFKPKYKFKFSQOY-----TAGRRL 455
1051 SKLECLINETSICENRKNELDOLKEAFKHEQEFITKLAFAEERNOMLLELETVOQAL 1110
456 SSLLTSHPEEYTKNGYLQKQVSISSD--ITSTYSSILRYTCGSSVEATRAV-KHLLA 511
1111 RSEMTDNONNSKSEAGLQKOEIMTLKEBONKMOKEVNDLOEN-----DOLMKVHK---- 1161
512 VYOHGCLGLSLAKRPLMRQESLOS--VNTEEOELIKAININSEVECGIHYLOIS-TSK 568
1162 -RKHEC-----QNISEPIINSVYERESERNOCFKPOMLEVEKISLSDGY 1206
569 SALSQFEAFQOGKSLIYNSGNIPDLPFFEFELPFCAS-ALDFIKLDYGGGMAWMEA 627
1207 NMOVLQELMLRKELKLOES-----EREKCLQHELOTITGDLTETSMLQ-MQ-- 1254
628 AEDTGGIHHBE--APETIYPS-RAVSLFFNMKQEPF-TLEVTLRDSKSKNKQOIT-LGKI 683
1255 SQEISLQKCEIDAEKRTISGPRELSTSONDANHLOCSIQITNNKLNLEKICEIHAERK 1314
684 FSSATSLRLQIRKAGVASSLSLVSTCKNIYSLMEVAPL--TIED----- 728
1315 YELVTELANRSRSCITATKRAKAEVGLNEVKITLMDSDGLLGELEVEDPGGEFSQEN 1374
729 ERHITSVTL-----KTLSHDLQNRPLPGGLTDSLGNKLNITKLIMD----- 771
1375 ECHPVSLAPLDESNSYHTLSLDRKEVOMHFAELQEKFLSLQSEHKILHOCOMSSIMBE 1434
772 -----NKKMNEDAIK-----LAEGL-KNLKKMCLFHLTHLSLDTIGEM 808
1435 LQTYVDSLAKENLVLTNLRNFGDLVKEMQGLGELGLVPSLSSSCVPDSSSLSSIGDSS 1494

809 DY-----IVKSLSEPCDLEIQLVSCCLSANAYK-----IIAQNLHN--- 846
1495 FYRALLEOTGDMSLLSNLEGAVASANQSCVDEVFSSLOITVYDLSLAENLVSTNLRNQG 1554
847 -LVKLSLIDLSENVL-----EKDGNALHELIDRNNVLEQITALLMP 887
1555 DLVKEMQGLGELGLVPSLSSSCVPDSSSLSSIGDSSFYRALLEOTGDMSLLSNLEGAVA 1614
888 WCGDVQSSLSLLKH-----LEEVQVLKGLKNNWRLTDEIRILGAFGKNPLKKNFQJN 943
1615 NQCSVDEVFSSLOENLTKRETTPAPAKGYE-----ELESICEYR-----QSLLE 1660
944 LAGNVSSDGMWLAFGVFNELKQLVF-----PDSFTEFLDPALVRKLSQVLSKLYELQ 998
1661 KLEEMMESQGIKKNEI-QELEQLLSERQELDKRQYLSN--EOWOOKITSVTEM 1716
999 EARLVGMCFFDDDLGV 1014
1717 ESKLAERKQTEQLSL 1732

RESULT 11
US-08-947-823-3
Sequence 3, Application US/08947823
Patent No. 6114605
GENERAL INFORMATION:
APPLICANT: Williamson, Valerie M.
APPLICANT: Kaloshian, Jsgouhl
APPLICANT: Vaghebbi, Jafar
APPLICANT: Bodessu, John
APPLICANT: Milligan, Stephen
TITLE OF INVENTION: Procedures and Materials for Confering
TITLE OF INVENTION: Pest Resistance in Plants
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-8834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/947,823
FILING DATE: 09-OCT-1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US97/18802
FILING DATE: 09-OCT-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/028,191
FILING DATE: 10-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Bastian, Kevin L.
REGISTRATION NUMBER: 34,774
REFERENCE/DOCKET NUMBER: 023070-070210US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1255 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: Protein
US-08-947-823-3

Tue Mar 26 11:36:23 2002

us-09-697-089-2.1a1

2.7%: score 144; DB 3; length 1255;
Query Match Similarity 18.1%; Pred. No. 0.00067; Indels 452; Gaps 56;
Best Local Similarity 169; Mismatches 358; KGSRC- 66
Matches 216; Conservative 169; Mismatches 358; KGSRC- 66

28 FVWVNLNEEVNICEVYDQARG---TTHMILK-----KGSRC- 66
191 FVWVNLNEEVNICEVYDQARG---TTHMILK-----KGSRC- 66
67 NLEKLS-----LKEWNPPLFODL-----NGSFLHOTSBDLDAOLKXLY 109
251 VGLFKOLLETSPDLIREVLPDHNHVVTPSTGARNHVMWELLLITLSDMPKFI 310
110 HPSPLNFY-----PLGSD-----DITFNKSTFEPVIM 140
311 HHDKLELDLVGLTREVSTYVRLDEEPNKEGNMOTNCATLIDLENE-LL 363
141 RKDQHNRYEOL-----LNGILOA-10SPCIEGSGKSTL 178
364 KKKLKHVYLAALDSSCCFPMSPDGLFMHLLHINLIDLSNAYSLAKELELYKODL 423
179 LORIAMGSGCKALTKRFKFEVFLSLRQGGIFETLCOULDI-----223
424 -----KFTIRSFV-----DABOGLYKDMARKVFLDGYNEKFPONCEIHALIKENHFKN 273
224 -----PGTIRKOTFAMMLKLRQVYFLDGYNEKFPONCEIHALIKENHFKN 509
464 NGHLHLIFSLPIIKK-----IKLKEISALD-----EMIPDRGLIYVNSPKRP 332
274 WVTYTTTECLRHIFGALFVAGDMTEPSAOLIREVLIKEALBGLL-1010KSRCL 551
510 VERKSITFDKI-----TVGEFEYN-----TLKRLKLSGADLVISITGAP 378
333 NLKPPLEFVVI-----TCALOMGSEFSPHSTOTLTFEPDILLOKRNKH 378
552 GSKTTLAVKYVNDKSVSRDLFRACVYDQGD-----KRLNLTISQVSDSKLS 605
379 KQVNASDFIRSLDHCGDLAEVFSHKEFDELQDVSVNEVLLTGLCKYTAQREK 438
606 ENIDVADKL-----KOLFGKRYLIVLDVYDWTTPWELT-----640
439 YKFFHKSROYTAGRSLSLTSHPEEYTKNGYLOK-----VSISDTSTYSILRYTGS 497
641 -----RPPEKSGSRI-----LITREKEVALHCKLNTDRLRLRDESMELLEKRAFGN 693
498 S-----VEATRAYMKH-----LAAYOHCLLGISIAKRLPMR-----QESLOAKNTEOE 544
694 ESCPDELIDVGEKELAEKGLPLVADLAGVIAORE-KKRSVWLEVOSSLSSTPLNSEVE 752
545 ILKAINISFVGCIIHLYOESISKALSOFEAFQOKSLYNSGNIPIYLFDFEHLPN 604
753 VMKAYELIS-----SALDFIKLDFYCGAMASWEKAEDTGTGIMEPAPETYPSPRAVS- 650
605 -----CA-----SALDFIKLDFYCGAMASWEKAEDTGTGIMEPAPETYPSPRAVS- 650
767 HLPDLTYFASPKDTSILTYELNYPFGABGFWGKTEMS-----NEEVKUYMDLITYSS 822
651 -----LEFN-----WKOEFRTLEVTLRDESKINKODITYLGIKITSANS-LRLOT-----694
823 LVICFNEIGYALFOIHDL-VHPCCLIKARKENLPDORSASDLDLPDIOITDCDEEH 881
695 -----KRCGVAGSLSVLSTCKNIYSIWEASPL-----TIDEDEHITSYTNK 740
882 FGLNFMEDSKNRHSG-----KHLYSLITIGDOLDSVSDAPHLHKLRLV 929
741 LSHDLONORLPGLTDSLOKLNKTLIMDINKMNEEDAIKLAGLKNLKKCLF-HIT 799
930 LQDH-----TSFLM-----VYDSL-----LNEICMLNHLR 954
800 HLDIGCMDYIYKSLSEPCDLEET-OLVSCSLAN-AVKILANLHNVKSLIDLSE 857
955 YLS-----IDVOKYL-----PLSFSNLMNLSLFTVSNRSILVLPRLIDLVKRLVSDA 1006
858 -WLEKDGNEA-----LHLELI-----DRMNVLPDLTALML-----PWGC 890

1007 CSFPMADSESLIADETKLENTLITELLISYKDKTKIKRPPNOLLSFEKESMDY 1066
891 DVQSGISLKLHKEEVPOLYKILKNMLD-----TEIRILGAFEGKN 934
1067 STQOHFSLDELETL-ETLSVGFKSNTNDSSGVATNRPWDFHPSKTLT-WUREF 1123
935 P-----LKNFOOLNAGNRVSSDCW-LAPWGVENLQOLVFPDFSKER 977
1124 PLTSDSLSTARLPNLELSLHTLHIGEEWNGEEDIFENIKELFNNOVSIKRW 1178

RESULT 12
US-09-180-439-6 Application US/09180439
Sequence 6, Patent No. 6225532
GENERAL INFORMATION:
APPLICANT: Hatzixanthidis, Kostas
APPLICANT: Hatzixanthidis, David A
APPLICANT: Jones, Jonathan DG
APPLICANT: Jones, Jonathan
TITLE OF INVENTION: Plant pathogen resistance genes and uses thereof
FILE REFERENCE: 620 - 53
CURRENT APPLICATION NUMBER: US/09/180.439
CURRENT FILING DATE: 1998-12-06
EARLIER FILING DATE: 1997-05-08
EARLIER APPLICATION NUMBER: GB 9609681.3
EARLIER FILING DATE: 1996-05-09
EARLIER APPLICATION NUMBER: GB 9619924.5
EARLIER FILING DATE: 1996-09-24
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 6
LENGTH: 799
TYPE: PRT
ORGANISM: Lycopersicon esculentum
US-09-180-439-6

2.7%: score 143.5; DB 4; length 799;
Query Match Similarity 23.7%; Pred. No. 0.00035; Indels 99; Gaps 17;
Best Local Similarity 62; Mismatches 128; SIREBIGYLS 312
Matches 258

675 QDITFKITSSATSLRLOKRCAGVAGSLVLTCKNIYSIWEASPL-TIDEERH 732
209 EETGTL-----RSLTKSLDIN-----FLSGSIPASISGNLNLSFYLYNOLSGSIPER-258
733 TSYTNKLTSHDLOKRLDGLTDSIGLKNLTKLIMDINKMNEEDAIKLAGLKNL 312
259 -----IGLRSITLYLDKKNALNSIPASISGNLNLSFYLYNOLSGSIPER-827
793 WCLEHHLHSDIG-----EGMDYIYKSLSEPCDLEIOL-----827
313 LTNLUGNNSLIGLIPASFGNNRNLQALFLNDNLLIOEIPFVCMWLSLELLYPRNNIK 372
828 -----VSCCL-----SANMYK-LAONHNLVKSITLISENTEKD-----GREAL 868
373 GKPVOCIGNISDLVLSMNSNFGELPSISNTLSLILDFGRNNLEGAIFQCGNITS 917
869 HELIDRANVLEOLTALMLPWGCG-----LEEVQLYKYLKRW 917
433 LOVFPMON-NKLSG-TLPNFSIGCSLISLHNGLEDELPMWLDCKKLYLDLGN 489
918 RLTDY-----EIRILGAFEGK-NPLK-----NFOOALNAGNRVSSDCW-LAP 546
490 QLNDTPMNLGTLPELRVLRITSKLHGPIRSSGAETMFPDLRITLDSNAPSOD-LP 546
958 MGVENLQOLVFPDFSTKE 976
547 TSLFELKGMKTYDKTME 565

Tue Mar 26 11:36:23 2002

us-09-697-089-2.ra1

Page 10

RESULT 13
US-08-514-014-4
Sequence 4, Application US/08514014
Patent No. 5707829
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John
APPLICANT: Carlin, Kerry
TITLE OF INVENTION: DNA SEQUENCES AND SECRETED PROTEINS
NUMBER OF INVENTION: ENCODED THEREBY
ADDRESS: 87 Cambridgepark Drive
City: Cambridge
State: Massachusetts
Country: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
CURRENT APPLICATION DATA: Patent Release #1.0, Version #1.25
APPLICATION NUMBER: US/08/514,014
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
TELEPHONE: (617) 498-8224
TELEFAX: (617) 498-8224
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 661 amino acids
TYPE: amino acid
MOLECULE TYPE: protein
US-08-514-014-4

Query Match
Best Local Similarity 2.7%; Score 142; DB 1; Length 661;
Matches 125; Conservative 65; Mismatches 215; Indels 140; Gaps 24;
486 TYSLLRYTCSSVEATRAVK--HIAAYOHGCLIGLSIAKRP--WROESLOSVKNT 541
DB 72 TTSRLMNT--FDLTRCOIMIHEDTPOSHQSLVLTGNPLIFMAETSLNGPKS-- 126
OY 542 BOELIKAININSVEGCIHLYOESTKSALSOEF--EAFPOGSKLYINGNITDYLPDF 598
DB 127 --LK--HFLIOTGISNL--EFLPVNLENLSLYLSNHI--SIKFE-- 166
OY 167 --FDEPPAENLKVDFOMNAHITSREDKRSLE--MAMSKA--AEDGGIIMTEAET 642
OY 643 YIPS--RAVSLFPNMKOFETLEVTRDSKLNKODIT--YLGKLFSSATSLRLQ 693
DB 223 VFOSLNGGPNLSVIFNGLONSTOSLWIGTFEIDDEBDSMMLGCLKEMSVETLNQ 282
OY 694 IKRCAGVAGSL--SIYSTCKNYSLSLIVAS 722
DB 283 EHFSDISITFCFOTLOELDLTANHLKLPSCMKGLNLTAVSY--NHFOI-QDIS 340
OY 723 PLTIDEHRTSVNKTSLIDNOLNORLPGCLTSLGMLKMLTGLINDINKMEEYALK 782
DB 341 AANPSTLTHYINGNKKL--HGVGCLERLNLDTLD--LSINDIEMSI CCS 389
OY 783 LAGELNKKMKCLPHLTILSDIGDMYIVKLSSEPC-DLEETOLVSCCLSAANVKTIA 841

DB 390 LQ--LNLSLHLOTNLISHNEPIG--LOSQAFKCEQLLELDLAFRLHINAFOSPF 441
OY 842 ONLHNVKLSIT--LDLSE--NYLEKQNGALHEILIDRMNVLEDTAL--M 885
DB 442 ONLHNVKLSIT--LDLSE--NYLEKQNGALHEILIDRMNVLEDTAL--M 885
OY 886 LPWGCVOGSSLSLKILEEVPOLYKLGKMRRLDTETRIEITLAFGKPKLNFQOLNLA 945
DB 502 ILSSC--GLSLIDQAFHSLKMSHVDLSHNSITLDSIDSLSHLKS--TYLNLA 551
OY 946 GNRVS 950
DB 552 ANSLN 556

RESULT 14
US-08-833-823-4
Sequence 4, Application US/08833823
Patent No. 5969093
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John
APPLICANT: Carlin, Kerry
TITLE OF INVENTION: DNA SEQUENCES AND SECRETED PROTEINS
NUMBER OF INVENTION: ENCODED THEREBY
ADDRESS: 87 Cambridgepark Drive
City: Cambridge
State: Massachusetts
Country: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
CURRENT APPLICATION DATA: Patent Release #1.0, Version #1.25
APPLICATION NUMBER: US/08/833,823
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
TELEPHONE: (617) 498-8224
TELEFAX: (617) 498-8224
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 661 amino acids
TYPE: amino acid
MOLECULE TYPE: protein
US-08-833-823-4

Query Match
Best Local Similarity 2.7%; Score 142; DB 2; Length 661;
Matches 125; Conservative 65; Mismatches 215; Indels 140; Gaps 24;
486 TYSLLRYTCSSVEATRAVK--HIAAYOHGCLIGLSIAKRP--WROESLOSVKNT 541
DB 72 TTSRLMNT--FDLTRCOIMIHEDTPOSHQSLVLTGNPLIFMAETSLNGPKS-- 126
OY 542 BOELIKAININSVEGCIHLYOESTKSALSOEF--EAFPOGSKLYINGNITDYLPDF 598

```

Db 127 -----LK-----HLFLITGIGISNL--EFIPVHNLNLESLYLGSNHISSIKF-- 166
QY 599 FEHLPCASALDEIKLDFEGGA-----MASWEKA-----AEDGGIHMEAPPT 642
Db 167 ----PKDFPARNLKYALDFONNMIHYISREDMSLEBOAILNLSFNNGNNKGIELGAFDST 222
QY 643 YTPS-----RAVSLEFNNKQEFRTLEVTLRDPFSKLNKODIT---YLGKIFSSATSLRLQ 693
Db 223 VQSLSNFGTGNLSYIFNGLQNSTGTSMLGTFEODIDDDISSAMKLGICEMSVESLNQ 282
QY 694 IKRCAGVAGSL-----SLVSTCKNITYSLMVEAS 722
Db 283 EHRFSDISSTFFQCFQTOLELDLTATHLKGLPSGMKGLMLKKIYLSV--NHFDQLCQIS 340
QY 723 PLTIEDERITSVNLKTLSTIHDLQNRRLPGGLTDSLSGLNKLTKLIMNIMKNEDAIK 782
Db 341 AANFSLTLYIRGNVKKI-----HLGVGCLLEKLGMLQTLTD--LSHNDIEASDCS 389
QY 783 LAEGLKLNKKMCLFHLTHLSDIGEGMDYIVKSLSEPC-DLEIOLVSCSLSANNAVKILA 841
Db 390 LG--LKNLSHLQTLNLSHNEPIG-----LQSAFKECQQLDLDAFTRLHIMAPQSPF 441
QY 842 QNLHNLVKIST---LDISE-----NYLEKQNEALHELIDRMNVLEQTLAL--M 885
Db 442 QMLHFLQVNLTYCFGLDTSNQHLLAGLPYLRHLNLEKGNHFDQGTITKTLNLQTVGSLLEV 501
QY 886 LHWGCDVQSGTSLSLKHLLEVPQYKLGKLNMRDLTDERIRLIGAFGKNPKLPQOLNIA 945
Db 502 ILSSC---GLSIDQOAFHSLGKMSHVDLSHNSLTGDSIDSLSHLKG-----LYLNLIA 551
QY 946 GNRVS 950
Db 552 ANSIN 556

```

RESULT 15
 US-08-519-547A-6
 : Sequence 6, Application US/08519547A
 : Patent No. 5994082

: GENERAL INFORMATION:
 : APPLICANT:
 : TITLE OF INVENTION: Proteins Essential for the Expression of
 : TITLE OF INVENTION: Vertebrate MHC Class II Genes, DNA Sequences Encoding Same
 : TITLE OF INVENTION: And Pharmaceutical Compositions
 : NUMBER OF SEQUENCES: 6
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: FISH & NEAVE
 : STREET: 1251 AVENUE OF THE AMERICAS
 : CITY: NEW YORK
 : STATE: NEW YORK
 : COUNTRY: U.S.A.
 : ZIP: 10020-1104
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: MS-DOS
 : SOFTWARE: WordPerfect 6.1
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/519, 547A
 : FILING DATE: 25-AUG-1995
 : CLASSIFICATION: 435
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: EP94113378.7
 : FILING DATE: 26-AUG-1994
 : ATTORNEY/AGENT INFORMATION:
 : NAME: HALEY, JAMES F.
 : REGISTRATION NUMBER: 27,794
 : REFERENCE/DOCKET NUMBER: YOS-11
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: 212-596-9000
 : TELEFAX: 212-596-9090
 : INFORMATION FOR SEQ ID NO: 6:
 : SEQUENCE CHARACTERISTICS:

```

: LENGTH: 1130 amino acids
: TYPE: amino acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEITICAL: NO
: ANTI-SENSE: NO
US-08-519-547A-6

```

Query Match 2.7%, Score 141.5; DB 2; Length 1130;
 Best Local Similarity 19.2%, Pred. No. 0.00095;
 Matches 168; Conservative 122; Mismatches 335; Indels 251; Gaps 35;

```

QY 151 QLTNLGLQALQSP-----CIEGESGKSKSLDRIAMLWSGCKALTKRFK 199
Db 391 QLAQGLAEVLLAAKHNRRPRETRVIAVLGAGQCKSWAGAVSRAMACGR--LPQYDF 447
QY 200 VFFL-----RLSRQGGLEFETCDQLDIPGTIRKQTFMAM-----LKLROBVLFL 247
Db 448 VFSVPCHLNRRGDAYG-----LQDLFLSLEPQVLVADEYFSHTLRPRDVLIL 498
QY 248 DGVNEFKQNPCEIEIBALIKENHREKNMVIYTTTECLRHIFQFALTAEGVDMTED--SA 305
Db 499 DAFEELEAOD-----GRHSTCGPAPAEPCSL 525
QY 306 QALIREVILKELAEQ---LLQIQKSRLRNMKT-PLFVYITCAIQNGES----EFHSH 357
Db 536 RQLLAGLQKCLLRKLTLLTARPRGRVQSLSKADALFELSGFSMEQAQAVMYRFPSS 585
QY 358 TQTLFHTFYDLDLQKNK-----HKHKGVAASDFIRSLDHGODLALAEVFSHKPFELQD 412
Db 586 GMTF--HODRALTLRKDRPLLSHSH---SPTLCRAQQLSEALLE----- 626
QY 413 VSSVNEVDVLTGTGLCKYT---AQRPKPKYKFFKNSFOEYTAGRRLSLL---TSH 462
Db 627 ---LGEDAKLPSTLGLVGLGRALDPSGALAEKLKAMELGRHRRQSTLQEDQFSA 683
QY 463 EPEEYTKNGVLOKKAVNISDITSTYSS--LRLRYCGSSVEATRAVVK-----HLA----- 510
Db 684 DVRTWAMAKGLVQHPRAAESELAEPSPFLQCFGLALMALSLGELKKELPOLYALALPRK 743
QY 511 -----AVYOH-----CCLGLSTAKRPLRROESLOVKWTTQDEITKA 548
Db 744 KRPYNNMLEGVPRLAGLIFQPPARCGLALGPSAASVDRKQVLAARYLKRLOPGLIRA 803
QY 549 ININSFEC-----GI--HLYOESTSK-----SALSOEFAPFGKSLYI 586
Db 804 ROLLELLCAHBAEBAEAGIWOHVQDELPGRLSFLGTRLTPPDHVLGKALAEAGQDFSLD 863
QY 587 NSGNLPDYLDFEFELPNCASAL-DFIKLDEVYGGAMASWEKAEDTGT--HMEAPETY 643
Db 864 RSTGI-----CPSGLGSLVGL---SCVTRFRAALSDTVALLMESTLQHGCTK 906
QY 644 IPSRAVSLEFNNKQEFRTLEVTLRDFSKLNKODITYIKLITSSATSLSLQIKRCAG--VA 701
Db 907 LQQAEEKF-----TIEFFRAKSLKDVEDLKL-----VDTQATRSSEPTA 948
QY 702 GSLSLVSTCKNITYSLMVEASPLTIED--ERHITSVNLKTLSTIHDLQNR---PGGLT 756
Db 949 GELPAVRDLKLLEFALGPGVSGQAPFKVLRLITLTAASSIQHDLDAISNKGIDEGVSQLS 1008
QY 757 DSLGKLNLTKLIMNITAMNEDAIKLAEGKLNKKMCLFHLTHLSDIGEGMDYIVSLS 816
Db 1009 ATEPOLKSELTNLNLSQNNITDLAGVAKLAELPSL----- 1042
QY 817 SEPCLEEIOLVSCLSANNAVKIILAQNLANLVKSIIDLSENVLEKQNEALHELIDRMN 876
Db 1043 ---AASLRLSLYNNCTICVGAESLARVLPDMVSLRWMDVQYNKTTAGAGQDLAASLRCP 1100
QY 877 VLEDTALMLPWGCDVQSGTSLSLKHLLEVPQYKVL 912
Db 1101 HVELTAM---WTPITPFSVO---EHLQOQDSRISL 1129

```

Tue Mar 26 11:36:23 2002

us-09-697-089-2.rai

Search completed: March 25, 2002, 10:55:40
Job time: 40 sec

XX MPI: 2001-308628/32.
DR N-PsDB; AAS03945.

PT Isolated caspase recruitment domain-12 polypeptide and nucleic acids
PR encoding them, useful for treating and diagnosing disorders associated
PT with abnormal apoptosis such as cancer, arthritis and Alzheimer's
PT disease -

XX
PS Claim 9; Fig 1; 93pp; English.

CC The sequence represents a human caspase recruitment domain 12 (CARD-12)
CC polypeptide. CARD domains are found in a number of proteins that transmit
CC signals that activate apoptosis and inflammatory pathways in response to
CC stress and other stimuli. Therefore, CARD-12 and its corresponding
CC nucleic acid may be used in treatment and diagnosis of patients suffering
CC from disorders associated with an abnormal level (an increase or a
CC decrease) of apoptotic cell death or abnormal activity of stress-related
CC pathways. The disorders include cancer, viral infections (e.g. caused by
CC poxviruses, adenoviruses), autoimmune disorders (e.g. systemic lupus
CC erythematosus, arthritis), neurological disorders (e.g. Alzheimer's
CC disease, amyotrophic lateral sclerosis), haematologic diseases (e.g.,
CC aplastic anaemia, myocardial infarction, stroke), inflammatory and immune
CC system disorders (e.g. Crohn's disease, insulin-dependent diabetes,
CC contact dermatitis, psoriasis, graft rejection), bacterial infections
CC (e.g. tuberculosis, leptomatous leprosy), ischemic and hypoxic brain
CC injury, kidney ischemia/reperfusion injury, excitotoxic brain damage,
CC acute bacterial meningitis and liver disease.

SQ Sequence 1024 AA;

Query Match 100.0%; Score 5281; DB 22; Length 1024;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1024; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

YY 1 MFPIKNSRALIORMGNTYIKOITDLEFYMNVLNRENNVICCEKYEDOAAGIIMILK 60
Db 1 mffiknsraliorgmntvlykqtdlilymvnlnreenvnlccexvegdaagllimilk 60

YY 61 KSESNCNLFLSKLENKNVLFODLNGSLFPHOTSSEGLDDLDQLDKDLNHTPSTFLNFYL 120
Db 61 ksesncnlflsklenknvlfodlngslfphotssegliddldqldkdlnhtpsflnfypl 120

YY 121 GEDIDIIENLKSTFEPEVLWRKDHHHREVDLTNLGLAALOSPCITIESEGSGKSTLLQ 180
Db 121 gedidiielnkstfepevlwrkdhhhrevedltnglgaalospctiesegsgkstelq 180

YY 181 RIAMIMGSKCKALKRFKRFFELRLSRAGGFELTCDLDDIPGIRKOTFMAMLIKLR 240
Db 181 riamimgskckalkrfkrffelrlsraggfeltecdlldipgirkotfmamlklkr 240

YY 241 QRVLFLDGVNFKPONCEIEALIKENRFRNMVVTWTTCRLRIHQFGALFAVGDM 300
Db 241 qrvlfldgvnfkponceiealikenrfrnmvvtwttcrlrihqfgaltaevgdm 300

YY 301 TEDSAQALIREVLAKELAEGLLDIQISRCRLNLKMTPLFVVITCAIQMGSEEFHSHTQT 360
Db 301 tedsaqalirevlakelaeglldiqisrcrlnlkmtpulfvvitcaiqmgseefhshtqt 360

YY 361 TLFHFHYDLLIOKKNHKHKGAVASDFIRSLDHCGDLAEGVSSHKFDLELDVSYSNEDV 420
Db 361 tlfhfhydlliokknhkhkgavasdfirsldhcgdlaegvsshkfdleldvssynedv 420

YY 421 LTTTGGLCYTVTORPKPKFFHKFSPOEYTAGRRSLTSLSHEPPEVTPGNGLYLOKMVSI 480
Db 421 ltttgglcytvorpkpkffhkfsfoeytagrrsltslsheppevtpgnglylokmvsi 480

YY 481 SDITTYSSLILVYTGGSSVEATRAWMKHIAAYOHGCLGLSILAARPLROESLOSVMKT 540
Db 481 sdittysslilvytggssveatrawmkhiaayohgclglasilaarplroeslosvmkt 540

YY 541 *TEQETLKAININSYEBCGIHLTYQESTSKALSQEDEFAGFGSKSLVYINSGNIDPLYLDFEFE 600

DB	Accession	Protein	Length
Db	541	tegel l k in n s t v e o g h h y g e s e k s a l s e f e a f g k y l n s g h p o l y f d f f e	600
Qy	601	HLPNCASALDEIKLDFYGGAMASWEKAEDTGGIHMEEAPETYIPSRAYSLEFFNMKEFR	660
Db	601	hlpncasaldfikidifygamaswekaedtgylhmeapeytipersavlsffnwkgfr	660
Qy	661	TLEWTLRPFSLKNQODITTYGKTRSSATSLRLOIKRACAGASLSLVSTCKNIYSIWE	720
Db	661	tlevlrlrfsklnqgditylqkifssatslrlqkrcagagslsivlstckniyslwe	720
Qy	721	ASPLTIEDERHITSVTNKLKLSHDLONOBLPGGLTDSGLNKLNTKLMDNIKNMEEDA	780
Db	721	aspltiiederhitsuvtlnklslshdlngnrlpggltdslynlknklklnmnikmeeda	780
Qy	781	IKLAEGKLNLKMKCLFHLTHLSDIGEGMDYIVKSLSEPPDLEIOLVSCCLSANAVKL	840
Db	781	iklaeglknlkkmclfnlthlsdigegmdyivkslssppdleeqlvscclsanavkl	840
Qy	841	AQNHLNLVYKSLIDLSSENYLEKSGNEALHLLIDRMVLEDTALMLPQMGDOGSISLL	900
Db	841	aqnlnhnlvklslidlssenylekdgnealhelidrmvleqltalmlpwgcdvqgsissll	900
Qy	901	KHLEEVPLVYLGKNNRLLPDTERTIRIGAFPGKNPLKNFOOLNAGNRSSOGMLAFMGV	960
Db	901	khleevpjlvylgknnrllpdtetririgafpgknplknfnqnljagnrssogmlafmgv	960
Qy	961	FENLKOLVFEDFSTKEELPPPALVRLKLSQVLSKRLTELOEARLVGWOPDDDLSVITGAER	1020
Db	961	fenkqlvlfedfstkeelfppalvrylksqvlslkrlteloearylvgwopdddlsvitgaek	1020
Qy	1021	LVTA 1024	
Db	1021	lvta 1024	
RESULT 2			
ID	AAU02881	standard; Protein; 1204 AA.	
XX	AC	AAU02881;	
XX	CT	12-SEP-2001 (first entry)	
XX	DE	Human caspase recruitment domain 12 (CARD-12) polypeptide #2.	
XX	DE	Caspase recruitment domain: CARD-12; apoptosis; stress-related pathway;	
KW	KW	cancer; viral infection; poxvirus; adenovirus; autoimmune disorder;	
KW	KW	systemic lupus erythematosus; arthritis; neurological disorder; stroke;	
KW	KW	Alzheimer's disease; amyotrophic lateral sclerosis; hematologic disease;	
KW	KW	aplastic anemia; myocarcardial infarction; inflammatory disorder;	
KW	KW	Crohn's disease; insulin-dependent diabetes; contact dermatitis;	
KW	KW	psoriasis; graft rejection; bacterial infection; lepromatous leprosy;	
KW	KW	tuberculosis; ischemic brain injury; hypoxic brain injury;	
KW	KW	kidney ischaemia; reperfusion injury; acute bacterial meningitis;	
KW	KW	excitotoxic brain damage; liver disease.	
OS	XX	Homo sapiens.	
PN	XX	WO200130971-A2.	
XX	PD	03-MAY-2001.	
XX	PF	26-OCT-2000; 2000WO-US29643.	
XX	PR	27-OCT-1999; 99US-0161822.	
XX	PA	(MILL-) MILLENNIUM PHARM INC.	
XX	PI	Bertin J, Robison KE;	
XX	WP	WP1: 2001-308628/32.	
DR	DR	N-PSDB; AAS03946.	

XX Isolated caspase recruitment domain-12 polypeptide and nucleic acids
PT encoding them, useful for treating and diagnosing disorders associated
PT with abnormal apoptosis such as cancer, arthritis and Alzheimer's
PT disease -

PS Disclosure; Fig 2; 93pp; English.

XX The sequence represents a human caspase recruitment domain 12 (CARD-12)
CC polypeptide. CARD domains are found in a number of proteins that transmit
CC signals that activate apoptosis and inflammatory pathways in response to
CC stress and other stimuli. Therefore, CARD-12 and its corresponding
CC nucleic acid may be used in treatment and diagnosis of patients suffering
CC from disorders associated with an abnormal level (an increase or a
CC decrease) of apoptotic cell death or abnormal activity of stress-related
CC pathways. The disorders include cancer, viral infections (e.g. caused by
CC poxviruses, adenoviruses), autoimmune disorders (e.g. systemic lupus
CC erythematosus, arthritis), neurological disorders (e.g. Alzheimer's
CC disease, amyotrophic lateral sclerosis), haematologic diseases (e.g.
CC aplastic anaemia, myocardial infarction, stroke), inflammatory and immune
CC system disorders (e.g. Crohn's disease, insulin-dependent diabetes,
CC contact dermatitis, psoriasis, graft rejection), bacterial infections
CC (e.g. tuberculosis, leprosy), infectious diseases (e.g. bacterial
CC injury, kidney ischaemia/reperfusion injury, excitotoxic brain damage,
CC acute bacterial meningitis and liver disease.

XX Sequence 1204 AA:

Query Match 98.7%; Score 5214; DB 22; Length 1204;
Best Local Similarity 98.1%; Pred. No. 0;
Matches 1014; Conservative 1; Mismatches 1; Indels 18; Gaps 1;

QY 1 MNFTKDNRAIQRGMGVINQITDDLFVWVNLNREEVNIICCEVEDDARGITHTLK 60
DB :|||||
DB 140 VNFIDKSRALIQMGMTVKIQTDDLFVWVNLNREEVNIICCEVEDDARGITHTLK 199
QY 61 KGSESCNLFELSKLEKNYPLFQDINGOS-----LFHQTSEGLDDLA 102
DB |||||||
DB 200 KGSESCNLFELSKLEKNYPLFQDINGOSFEETQWVFNISLIGTHQSEGLDDLA 259
QY 103 QDLKDLVHTPSFLNYPFGEIDIDIIENLKSTFTEPVLMRKDHHHRYQLTLNGLQALQ 162
DB |||||||
DB 260 QDLKDLVHTPSFLNYPFGEIDIDIIENLKSTFTEPVLMRKDHHHRYQLTLNGLQALQ 319
QY 163 SPCTIEGSSGKSTILQRIAMIMSGCKALTKFKFPIRLISAOGLEFETLCDQLD 222
DB |||||||
DB 320 SPCTIEGSSGKSTILQRIAMIMSGCKALTKFKFPIRLISAOGLEFETLCDQLD 379
QY 223 IPGTIRKOTFAMMLKLRQVLFLLDGNENKPPONCEIEALIKENHFKNMVITTTTE 282
DB |||||||
DB 380 IPGTIRKOTFAMMLKLRQVLFLLDGNENKPPONCEIEALIKENHFKNMVITTTTE 439
QY 283 CLRHIFQFALTAEVGMTEDSAQALIREVLKELAEGLLQIOKSRCLNLMKTPLEFV 342
DB |||||||
DB 440 CLRHIFQFALTAEVGMTEDSAQALIREVLKELAEGLLQIOKSRCLNLMKTPLEFV 499
QY 343 ITCAIOWGSEFHSHTQTFTHFTFYDLLIOKNKHKKVAAASDFRSLDHGDLALBEVF 402
DB |||||||
DB 500 ITCAIOWGSEFHSHTQTFTHFTFYDLLIOKNKHKKVAAASDFRSLDHGDLALBEVF 559
QY 403 SHKDFPELDQVSVNEDVLTGGLCKTAAQRFKRYFFKFSFOEYTAGRRLSLTSH 462
DB |||||||
DB 560 SHKDFPELDQVSVNEDVLTGGLCKTAAQRFKRYFFKFSFOEYTAGRRLSLTSH 619
QY 463 EPEEVTGNGYLVQKMSISDITSTYSSILRYTCSSVEATRAVMKHLAAVYOHGCLGLS 522
DB |||||||
DB 620 EPEEVTGNGYLVQKMSISDITSTYSSILRYTCSSVEATRAVMKHLAAVYOHGCLGLS 679
QY 523 IAKRPIMROESLVKNTTEDEILKATININSFVGTHLYQESKSLSGEFAPFGK 582
DB |||||||
DB 680 IAKRPIMROESLVKNTTEDEILKATININSFVGTHLYQESKSLSGEFAPFGK 739

QY 583 SLVNSGNIPDYLFDFEHLPCASALDEIKLDFYGGAMASWEKAEDDGIHMEAPET 642
DB |||||||
DB 740 SLVNSGNIPDYLFDFEHLPCASALDEIKLDFYGGAMASWEKAEDDGIHMEAPET 799
QY 643 YIPSAVSLFFNMKQDFRLEVTLRDPSKLNKQDITTYLGKIFSSATSLRLIKRCAGVAG 702
DB |||||||
DB 800 YIPSAVSLFFNMKQDFRLEVTLRDPSKLNKQDITTYLGKIFSSATSLRLIKRCAGVAG 859
QY 703 SISLVLSCKNITSLMVEASPLTIEDERHITSVYNLKLTLSDHLDONQPLPGDLSGLNL 762
DB |||||||
DB 860 SISLVLSCKNITSLMVEASPLTIEDERHITSVYNLKLTLSDHLDONQPLPGDLSGLNL 919
QY 763 KNLTLINDNIRKMNEDIKLAEGIKNKKMCLPHLTLSIDGEGMDYIVKSLSEPCDL 822
DB |||||||
DB 920 KNLTLINDNIRKMNEDIKLAEGIKNKKMCLPHLTLSIDGEGMDYIVKSLSEPCDL 979
QY 823 EIIQIVSCCLSANAVKILIAQNLHNLVLSITDLSENYLEKGNENALHILIRMNVEQLT 882
DB |||||||
DB 980 EIIQIVSCCLSANAVKILIAQNLHNLVLSITDLSENYLEKGNENALHILIRMNVEQLT 1039
QY 883 ALMLPMGCDVQGSLSLKLHEVPQVLKLGKMKRLTDTETIRILGAFEGKNPLKNFOOL 942
DB |||||||
DB 1040 ALMLPMGCDVQGSLSLKLHEVPQVLKLGKMKRLTDTETIRILGAFEGKNPLKNFOOL 1099
QY 943 NLAGNRVSSDGLAFMGVEENLKOLVFPDFSTKEFLPPALVVRKLSOVLSTFLQEARL 1002
DB |||||||
DB 1100 NLAGNRVSSDGLAFMGVEENLKOLVFPDFSTKEFLPPALVVRKLSOVLSTFLQEARL 1159
QY 1003 VGMQFDDDDLSVIT 1016
DB |||||||
DB 1160 VGMQFDDDDLSVIT 1173

RESULT 3
AAM23595
ID AAM23595 standard; Protein; 841 AA.
XX
AC AAM23595;
XX
DT 12-OCT-2001 (first entry)
XX
DE Murine EST encoded protein SEQ ID NO: 1120.
XX
KW Human: sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder;
KW biodiversity; gene therapy; nutrition.
XX
OS Mus musculus.
XX
PN WO200154477-A2.
XX
PD 02-AUG-2001.
XX
PE 25-JAN-2001; 2001WO-US02687.
XX
PR 25-JAN-2000; 2000US-0491404.
PR 17-JUL-2000; 2000US-0617746.
PR 03-AUG-2000; 2000US-0631451.
PR 15-SEP-2000; 2000US-0663870.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI Cao Y, Drmanac RA, Zhang J, Werhman T;
XX
XX WPI: 2001-476164/51.
XX N-PSDB; AAH98254.
PT Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use -
XX
PS Claim 20; Page 832-833; 1275pp; English.

Db 799 pstkagpkivshllhvdn-----keslenisenddylkhpelslqmj 843
 QY 546 LKAI---NINSFVEGCIHL-----YOEYTSKSLAQEFAAFQGSYINSGNIPD 593
 Db 844 lrglwgicpqayfemwsehlvialktaqysnt-vaacspfvigfignrtltlganl-q 901
 QY 594 YLPD-----FPEHL-PNC--ASALDFIKLDFYGG--AMASWE 625
 Db 902 yfifhpeslslrsihfrlgnktsprahsvletcdksqvpfldqdyasafemewe 961
 QY 626 -----KAEDTGTGIMEAPETY-IPSRASVLPFNMKQEFRTLEVTL 666
 Db 962 rnlakeednvksymdngrtraspdlstgwklspxkyp-----cleadv 1006
 QY 667 RDESKLNKODITYLGKIFSSATSRLQIKRCAGVAGSLSVLSTCK-NIYSLWEASPLT 725
 Db 1007 ndldvvgqdmlelmvfassqrlhlnhsrgfiesirpalelsksavckcsklskls 1066
 QY 726 IEDERHITSVYNLKTLSIH---DLQNRPLPGGLTDSLGKLNKLTLMNDIKM----- 775
 Db 1067 aaeqellltplsleslsvsgtlsgdqdlfpn--ldkfclckelsvdlgnlnvfsvipee 1124
 QY 776 -----NEEDATIKLAEGLKLNKMKCLFHL--THUSDIGEMDYIVKSLSE 818
 Db 1125 ffnfhmeklllqisaeysdsklvklignspnlvfhkcnfisdqslmcmvls----- 1179
 QY 819 PC-DLEBIQIVSCCLSANAVKILQNLNLVLTSLDL-SENYLEKQNEALHELIDRMN 876
 Db 1180 -ckltsktsdftf--qavpfa-slpnflsklhlngqgfpdeetsekfayllsgls 1235
 QY 877 VLEDTLMLPMWCGDVGSLSSLLKLEBPQVLKLGKMKRLTDT-----EIRIICAF 931
 Db 1236 nlee--lllptgqlyrvaklllqgcqqlhcltrvlstfklindsvvelakvaissgf- 1291
 QY 932 GKRNKLNKNOOLNAGN-RVSSDGLAFMGVFNKOLVFPDFS--TKREPLDPAVLK 987
 Db 1292 -----qlenlkslnhklieegrynfqaldmnpnlqeldisrhfecikagatlvksl 1346
 QY 988 SQVLSKLTLFLOEARLVGMQFDDDLVAVI 1015
 Db 1347 sqcvlrlprlrlmlslwlladdatall 1374

RESULT 6
 ID AAY14079 standard; Protein: 1403 AA.
 XX AAY14079;
 AC AAY14079;
 DT 20-JUL-1999 (first entry)
 XX
 DE Gonadotropic hormone protein sequence.
 XX Gonadotropic hormone: excessive ovulation animal; transgenic animal;
 XX totipotent cell; somatic cell chromosome.
 OS Homo sapiens.
 XX JP1113444-A.
 PN JP1113444-A.
 XX
 PD 27-APR-1999.
 XX
 PF 14-OCT-1997; 97JP-0280830.
 XX
 PR 14-OCT-1997; 97JP-0280830.
 XX
 PA (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
 PA (SAKA/) SAKAI H.
 XX WPI; 1999-320709/27.
 DR N-PSDB; AAX58000.
 XX
 PT An excessive ovulation animal - useful for improving the

PT productivity of animals
 XX
 PS Claim 3: Page 11-14; 18pp; Japanese.
 XX
 CC This sequence represents a gonadotropic hormone.
 CC The invention relates to an excessive ovulation animal, which is a
 CC transgenic animal with a totipotent cell containing a DNA fragment
 CC containing a promoter sequence and a gonadotropic hormone coding
 CC sequence. The DNA fragment is in the somatic cell chromosome. The
 CC excessive ovulation animal is useful for improving the productivity of
 CC animals. The method can improve the productivity of a useful animal.
 XX
 SQ Sequence 1403 AA:
 Query Match 12.4%; Score 653; DB 20; Length 1403;
 Best Local Similarity 23.7%; Pred. No. 9, 1e-46;
 Matches 263; Conservative 197; Mismatches 410; Indels 236; Gaps 42;
 QY 69 FLKSLKEMVPLFODLNGQ-----SLPHQTSQGLDD----- 100
 Db 344 flqumks-saeavrpdlqsrgelettsenledslavgrlpvmagaeqwfgeakn 402
 QY 101 LAODKDLVTPPSFLN-----YPLGEDDIITNLSKSTFEVPLMKKDOHNR 146
 Db 403 lneqlraaytsasfrhmslldssdlatdhlldgdlsl---aaskhskpy----- 449
 QY 149 VEOLTLNGLQALQSPCTIGESGSKSTLQRLAMLMGSKCALTKFVFFLRSL-- 206
 Db 450 geplvlpvfnghlnsvmcveagsgkvtllkklafawascoplntfqlvlylsist 509
 QY 207 RAQGLFETLCDOLLDPGTIRKOTFAMLLKLQRYVFLLDGNEF--KQNCPEIEAL 264
 Db 510 rpedglaslcdqllekgsvtemcmrnllyqknglfllddykelcsipq---vlgkl 566
 QY 265 IKENHRKKNWVITTTTECLRHTRQEGALTAEVGDMEDBSAQLIRVILKELA--EGLL 322
 Db 567 lqkhnstlcllavlfnrardlrrylellelkafrlytvcrltkrlshnmtrlrlkim 626
 QY 323 LQIQSRCLNLMKTPLEVVITCAIQGESEFHSHTQTLEFHTFYDLLIQNKKHKNQVA 382
 Db 627 vyfgkngslqkqkprlfaivaicahwfgypfdpsfdvavfkymelstlrnk-----a 680
 QY 383 ASDFIR-SLDHCGDLAIEGVFSHKFDELDV--SSVNEVDVLTGILCKYTAQREPKY 439
 Db 681 taellkatvsscgealqkfscfendddlaeaydededltmcmsftqgrlrfy 740
 QY 440 KFFHKSQETAGRRSLSLTSHPEVYTKNGYLOKMSISDITSYSSSLRYTCSSV 499
 Db 741 rflspatqetlagnrlllellsdrgqbqdlgylhklqinspmmtvsaynflnyv--ssl 798
 QY 500 EATRA---VMKHLAAVYQHGCLLGLSLAKRPLRMWQESLQSVKN-----TTEQET 545
 Db 799 pstkagpkivshllhvdn-----keslenisenddylkhpelslqmj 843
 QY 546 LKAI---NINSFVEGCIHL-----YOEYTSKSLAQEFAAFQGSYINSGNIPD 593
 Db 844 lrglwgicpqayfemwsehlvialktaqysnt-vaacspfvigfignrtltlganl-q 901
 QY 594 YLPD-----FPEHL-PNC--ASALDFIKLDFYGG--AMASWE 625
 Db 902 yfifhpeslslrsihfrlgnktsprahsvletcdksqvpfldqdyasafemewe 961
 QY 626 -----KAEDTGTGIMEAPETY-IPSRASVLPFNMKQEFRTLEVTL 666
 Db 962 rnlakeednvksymdngrtraspdlstgwklspxkyp-----cleadv 1006
 QY 667 RDESKLNKODITYLGKIFSSATSRLQIKRCAGVAGSLSVLSTCK-NIYSLWEASPLT 725
 Db 1007 ndldvvgqdmlelmvfassqrlhlnhsrgfiesirpalelsksavckcsklskls 1066
 QY 726 IEDERHITSVYNLKTLSIH---DLQNRPLPGGLTDSLGKLNKLTLMNDIKM----- 775

Db 1067 aaeqelliltplsleeslevsgtigsqgdlfrn--ldkflclkelsvdleglnvfsvlpee 1124
QY 776 -----NEEDAIKLAEGLNKKMCLFHL--THLSDIGCMYIVKSLSE 818
Db 1125 ffnfhmeklilqisaeydpsklvklqnsplnhvfhkcnffsfqsglmmlvs----- 1179
QY 819 PC-DLEEIOLVSCCSAANAIVKILAOINLHVKSLTLD--SENYLEKDCNEALHELIDRMN 876
Db 1180 -cklteikfsdfff-qavpfva-slpnfisiklinlegqgfpoeeekfayllgsls 1235
QY 877 VLEQLTALMLPWGCDVGSLSLKLHLEEVDPQVYKLGKNNRLDPT---EIRILGAF 931
Db 1236 nlee---ilptgdlgyrvaklilqgcqqlnclyvlsffkclndssvveakvaissgff- 1291
QY 932 GKNPLKKNQOQLNAGN--RVSSDGMWLAFMGVFENLKOYVFPDES---TKEFLPDPALVAKL 987
Db 1292 -----qklenklkslnhklteegynrfqaldmnpnlgelidstrftecikagattvksl 1346
QY 988 SQVLSKLTFLQEARLVGQWQFDDDLSTV 1015
Db 1347 sqcvlrlprlrlrlmmlswldaddiall 1374
RESULT 7
AA09539
ID AA09539 standard; Protein; 1403 AA.
XX
AC AA09539;
XX
DT 20-JUL-1999 (first entry)
XX
DE Human apoptosis inhibiting protein #1.
XX
KW Human; apoptosis inhibitory protein; apoptotic disease; diagnosis;
KM spinal muscular atrophy.
XX
OS Homo sapiens.
XX
PN JP1116599-A.
XX
PD 27-APR-1999.
XX
PF 14-OCT-1997; 97JP-0280831.
XX
PE 14-OCT-1997; 97JP-0280831.
XX
PS 14-OCT-1997; 97JP-0280831.
XX
PT (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
XX
DR WPI: 1999-323531/27.
XX
N-PSDB; AAX56272.
XX
PT New apoptosis inhibitory protein - useful for determining mechanism
PT of various apoptotic diseases e.g. human spinal muscular atrophy
XX
PS Claim 1; Page 4-8; 16pp; Japanese.
XX
CC The present sequence represents a human apoptosis inhibitory protein.
CC The apoptosis inhibitory protein is useful for the elucidation of
CC the mechanism of various apoptosis diseases such as human spinal
CC muscular atrophy and the diagnosis, the prevention and the treatment
CC of such diseases.
XX
SQ Sequence 1403 AA:
Query Match 12.4%; Score 653; DB 20; Length 1403;
Best Local Similarity 23.7%; Pred. No. 9,1e-46;
Matches 263; Conservative 197; Mismatches 410; Indels 238; Gaps 42;
QY 69 FLKSLKENNYLPFDLNGQ---SLFHOTSGDLDL----- 100
Db 344 flqmkss-saevtpdlgrgclcellettsesnldestiavgiyvpemagagwfgaekn 402

QY 101 LAODCKOLYHPSFLNF-----YPLGEDIDILFNLKSTFEPVLMRKQDHNHR 148
Db 403 lneqrlraaytsesfihmslaldissldatdhllgdcldsl-----askhskpv----- 449
QY 149 VEOLTLNGLLOALOSPCTIEGESGKSTLQRIAMLMGSGKRALTFKFEVFLRLS-- 206
Db 450 qeplvpfvgfnlsvmevegeagsgkvlkklafvwasgcprrlnfqlvfytsbst 509
QY 207 RAOGGLFETLCDQLDITGTRIKOTFMAMLLKROVRVFLDUGYEF--KRONCPETAL 264
Db 510 rpedglaslicdqlllekegsyvecmemrlilqqlknqvlfllddykelosbpq---vialk 566
QY 265 IKENIRFRKNMVVTFTTCLRIHROFALTAEVGOMTEDSAQALREVLIKELA--EGLL 322
Db 567 lqknhlstrclliavtrnraditrryletlleikafpynvcllrlkflshmmlltrkf 626
QY 323 LOIQSRCLRLMLKTPLEFVITCAIOMGESEFHSHTQTLTFHTFYDLLQKNKHKQVA 382
Db 627 vyfgkngslqkqktrplvvaicahwfgypfdpsfdavafksymernsltnk-----a 680
QY 383 ASDPFR-SLDHCGDLALGCVSHKDFELQDY--SSVNEVDVLLTGGLCKYTAQRFKRY 439
Db 681 laelkavtsscgelalqkfiscfcdfeinddlaeagvdeddlmclnlskfvaqrlrpfy 740
QY 440 KEFHKSFOEYTAGRRLSILTSHEPEVTKGNGYLOKKWVSTDIYSTVSSLLRYTCSSV 499
Db 741 rflspatqelflagmrllellsdqehqdlqlyhklqjnspmmtvsaaynflnyv--ssl 798
QY 500 EATRA---VMKHLAAVYOHGCLLGLSIARKPLRMOESLOSRYN-----TTEOEI 545
Db 799 pstkegpkvlshllhlvdn-----keslenlenddyllkhpelslqql 843
QY 546 LKAI-----NINSEVDCGTHL-----YQESTKSALSQEFKFAFGCKSLYINSGNIPD 593
Db 844 lrglwgqicpqayfsmvsehllvlalktaysnt-vaacspflqlfgrtrtlialn-l-q 901
QY 594 YLPD-----FPEHLPMC--ASALDFIKLDFYGG--AMASWE 625
Db 902 yffdhpeslslrlsrhfrignktsprahfsvlctcfksqyrltdqdyasafermewe 961
QY 626 -----KAAEDTGGIMEEAPERY-IPSRAVSLFFMWKQEFFRLEYTL 666
Db 962 rnlackednvksymdmqrraapdsistgyklspkqykfp-----cleadv 1006
QY 667 RDFSRLNKQDITTYLAKIRSSATSLRLQIKCAVAGSLSLVSTCK-NIYSLMVEASPLT 725
Db 1007 ndldvvgqdmlelmntvfaaqrlhlnhargfieslrpalelskavtksisklels 1066
QY 726 IEDERHITSVNLKTLSTH---DLONORLPGGLPDSLONLKNLTKLINDNKM----- 775
Db 1067 aaeqelliltplsleeslevsgtigsqgdlfrn--ldkflclkelsvdleglnvfsvlpee 1124
QY 776 -----NEEDAIKLAEGLNKKMCLFHL--THLSDIGCMYIVKSLSE 818
Db 1125 ffnfhmeklilqisaeydpsklvklqnsplnhvfhkcnffsfqsglmmlvs----- 1179
QY 819 PC-DLEEIOLVSCCSAANAIVKILAOINLHVKSLTLD--SENYLEKDCNEALHELIDRMN 876
Db 1180 -cklteikfsdfff-qavpfva-slpnfisiklinlegqgfpoeeekfayllgsls 1235
QY 877 VLEQLTALMLPWGCDVGSLSLKLHLEEVDPQVYKLGKNNRLDPT---EIRILGAF 931
Db 1236 nlee---ilptgdlgyrvaklilqgcqqlnclyvlsffkclndssvveakvaissgff- 1291
QY 932 GKNPLKKNQOQLNAGN--RVSSDGMWLAFMGVFENLKOYVFPDES---TKEFLPDPALVAKL 987
Db 1292 -----qklenklkslnhklteegynrfqaldmnpnlgelidstrftecikagattvksl 1346
QY 988 SQVLSKLTFLQEARLVGQWQFDDDLSTV 1015
Db 1347 sqcvlrlprlrlrlmmlswldaddiall 1374

RESULT 8
AA88053
ID AAY88053 standard; Protein: 1403 AA.
XX
AC AAY88053;
XX
DT 22-SEP-2000 (first entry)
XX
DE Human NAIP protein.
XX
KW NAIP: apoptosis inhibiting protein; monoclonal antibody; diagnosis;
KW apoptosis disease onset mechanism; drugs development; prevention;
KW treatment; human.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Region 256..586
FT /note="Immunogenic region, specifically described
FT in Claim 1"
FT Region 841..1052
FT /note="Immunogenic region, specifically described
FT in Claim 1"
XX
WO200024889-A1.
XX
PD 04-MAY-2000.
XX
PF 22-OCT-1999; 99WO-JP05841.
XX
PR 26-OCT-1998; 98JP-0304550.
XX
PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.
XX
PI Sakai H, Ikeda J;
XX
DR WPI: 2000-350731/30.
XX
DR N-PSDB; AAA39808.
XX
PT Simple and accurate quantitation of human apoptosis inhibitory protein
PT (NAIP) with monoclonal antibodies, for disease diagnosis and
PT development of drugs to prevent and treat apoptosis-related diseases
XX
PS Claim 1: Page 24-30; 36pp; Japanese.
XX
CC This invention describes novel anti-NAIP monoclonal antibodies, produced
CC by hybridomas, prepared by fusing antibody-producing cells of mammals
CC immunized with an immunogen containing residues 256-586 or 841-1052 of a
CC 1403 residue amino acid sequence, fully defined in the specification, or
CC their partial sequence, with a myeloma cell line. The monoclonal
CC antibodies are for the study of apoptosis disease onset mechanism,
CC disease diagnosis and development of drugs to prevent and treat
CC apoptosis-related diseases. The quantitation method with the monoclonal
CC antibodies is simple and accurate by using a biological specimen. This
CC sequence represents the human apoptosis inhibitory protein NAIP which is
CC described in the method of the invention.
XX
SQ Sequence 1403 AA;

Query Match 12.4%; Score 653; DB 21; Length 1403;
Best Local Similarity 23.7%; Pred. No. 9, 1e-46;
Matches 263; Conservative 197; Mismatches 410; Indels 238; Gaps 42;

QY 69 ELKSLKEMNYPLFQDINGO-----SLFHQTSQDLD----- 100
DB 344 fllgmks-saeyrpdqrsrgelcelllettsesnlavgrlpvemaqgeaqwfgeakn 402
QY 101 LAQDLKDLHTPSFLNF-----YPLGEDIDIFNLKSTTEVYLMRKDQHNHR 148
DB 403 lneqlraaytsasfrmslldssatclatclhllgdcldsl--askhiskpv----- 449

QY 149 VEQTLNGLQALQSPCIIIEGSGKSKSTLQRIAMLMWGSCKALTKFVFEFLRLS-- 206
DB 450 qeplvlpewfgnlnswmcvegeasgkvtllkkaifawscgcpnlrftglvylslet 509
QY 207 RAQGLFETLQDQLDIPGTRKQTMAMILKTRQVLPFLDGYNEF--KPNCPETAL 264
DB 510 rpedglasllcdqlllekgsvtemcmtnllqllknvllfllddykeicslpq--vlgk 566
QY 265 IKENHRFKMNVITTTTECRHTRQFALTAEVGDMTQEDSAQALREVLIKELA--EGLL 322
DB 567 lqknhlstrcllavtrnarrdiryletlleikafryntvcilrklfshmttrllrk 626
QY 323 LQIQSRCLRLMKTPDLFVVTTCALQMGSESEFHSHTQTLFHTFYDLLQKNKHKGVA 382
DB 627 vyfgkngsqkqkqkplftraalcahwfyqfdfsfdavafksymerslrlrk-----a 680
QY 383 ASDGFR-SLDHCGDLALLEGVFSHKDFPELODV--SSVNEDEVLTTLCKYTAQRKPKY 439
DB 681 taelikavsscgelaikgffscfcfeinddlaaagvdeddlumcmekftaqrlrpfy 740
QY 440 KPEHKSFOEYTAGRLISLLTSHPEEVTKNGXYLQKMYSIDITSTYSLLRITGSSV 499
DB 741 rflspafgefiaqrlllelidsrgehqdgllqylhqlqinspmntvasaynflny--ssl 798
QY 500 EATRA---VMKHLAAVYOHGCLGLSIARKRPLWROESLOSVKN-----TTEOEI 545
DB 799 pskagapkllvshllhvdn-----keslenisenddyllkqpeislqmql 843
QY 546 LKAI---NINSFVECGIHL-----YQESTSKALSOEFAPFGOKSLYINSGNIPD 593
DB 844 lrlglwqicpqayfsmvsehllvalklaygnl-vaacspfyqlfqlgrtlrlglaql 901
QY 594 YLFD-----FPEHLPNQ--ASALDFIKLDPYGG--AMASWE 625
DB 902 yfildpesislrlrinfirngktsprahfsvlelctfdksqypltdqdasafepmewe 961
QY 626 -----KAEDTGCIHMEAPERY-IPSRVSLFPMNKQERTLEVTL 666
DB 962 rlaekednvkymdmqgraspsldstgykklspkqklp-----cleadv 1006
QY 667 RPSKLNKODIYTLGKIFSSATSRLQIRKACVAGSLSVLSTCK-NIYSLMVEASPLT 725
DB 1007 ndldvvgqdmlellmtvfaasqrllehlhnsrgfiesirpalelskavtkcsisklels 1066
QY 726 IEDERHITSVTNLKTLTSLH---DLQNRQLPGGLTOSLGNLKNLTKLIMNIMK----- 775
DB 1067 aaeqellilcplslesvsgtlsgdqifpn--ldkflclkslvadlegnlnvfvipee 1124
QY 776 -----NEEDAIRKLAEGIKMLKMKCLFHL--THLSDIGEMDYTVKSLSE 818
DB 1125 fndfhmeklllgiseeypskrlkllqnsplnhvfhlkcnfstdsglmtlvs----- 1179
QY 819 PC-DLEFIQVSCCLSANVAKILAQNLHLVKSITLD--SENYLEKDGKALHELIDRNN 876
DB 1180 -ckltekfsdsff--qavpfva-slpnflskllnlegqfpdeelsekftayllgslls 1235
QY 877 VLEQTLALMLPMPCDVGQSSLSLKLHEEVPOLYVKGKLMWRLLTPD-----EIRIIGAF 931
DB 1236 nlee---lllpgdgilyraklllqqcgqlhclrylsfktlnddsveiaakalsgfl 1291
QY 932 GKNPLKPNQOLNLAGN-RVSSDGLWLAFMGVFNELKOLVFEFDS---TKFELDPALVRL 987
DB 1292 -----qklenlkslnhktleegyrnffgaalnmplqglldlsrhtecikagatvksel 1346
QY 988 SOYLSKLTFLQEARLYGVGFDDDLSTV 1015
DB 1347 sqcvlrlprlrlrlnmlswlldaddall 1374
RESULT 9
AAW20033
ID AAW20033 standard; Protein: 1403 AA.
XX

AC AAM20033;
XX
XX 06-OCT-1997 (first entry)
XX
XX
DE Neuronal apoptosis inhibitor protein (NAIP).
XX
XX Neuronal apoptosis inhibitor protein; NAIP; diagnosis;
KM therapy; cancer; AIDS; amyotrophic lateral sclerosis;
XX spinal muscular atrophy.
XX
OS Homo sapiens.
XX
XX
FH Key Location/Qualifiers
FT Misc-difference 600 /note= "encoded by AAA"
TT Misc-difference 919 /note= "encoded by CCA"
XX
XX
XX MO9726331-A2.
XX
XX 24-JUL-1997.
XX
XX 17-JAN-1997; 97WO-IB00142.
XX
XX 19-JAN-1996; 96GB-0001108.
XX
XX (UYOT-) UNIV OTTAWA.
XX
XX korneluk RG, Mackenzie AE, Robertson G, Roy N, Tamai K;
XX WPI: 1997-385335/35.
XX DR N-PSDB; AAT71266.
XX
XX
XX New neuronal inhibitor of apoptosis - useful for diagnosing and
PT treating, e.g. cancer, AIDS or amyotrophic lateral sclerosis
XX
XX
PS Claim 41; Fig 7A-L; 102pp; English.
XX
XX
XX Novel human neuronal apoptosis inhibitor protein (AAM20033), or NAIP,
CC is a negative regulator of apoptosis. Partic. neuronal apoptosis
CC and, when deficient or absent, contributes to neurodegenerative
CC phenotypes such as spinal muscular atrophy (SMA) and amyotrophic
CC lateral sclerosis. Its amino acid sequence was deduced from a
CC cDNA clone (AAT71266) obtd. from a human foetal spinal cord cDNA
CC library. NAIP polypeptides, esp. those containing at least two
CC BIR (baculovirus IAP repeat) domains, can be expressed in host-
CC vector systems and used to increase or induce apoptosis for the
CC treatment of AIDS, neurodegenerative disease, myelodysplastic
CC syndromes or ischaemic injury, to screen for (ant)agonists, or to
CC produce antibodies useful for inhibiting apoptosis.
XX
XX
XX Sequence 1403 AA;
SO

Query Match 12.3%; Score 652; DB 18; Length 1403;
Best Local Similarity 23.7%; Pred. No. 1.le-45;
Matches 261; Conservative 196; Mismatches 418; Indels 228; Gaps 41;

QY 69 FLKSLKENVYPLFDLNGO-----SLFPHOTSGDLDD----- 100
DB 344 flgmks-saevtpdlrgelcellettseslnedsiaavpivemagseqawfgeakn 402
QY 101 LAODLKDYHPPSFLNF-----YPLGEDIDIFNLKSFTEPVLVRKQDHHNR 148
DB 403 lneqlraaytsstfrimeldsldatclhllgdisl---askhskrv----- 449
QY 149 VEQLTLNGLAIOGPCIIEGSGKSTLQRIAMLMWSGCKALTEFKFVFLRLS-- 206
DB 450 qeplvlpvfgnlnsvmcvegeagsgkvllkklafwaagccpllnfqlvfyislst 509
QY 207 RAQGLFLETLCDQLDIDIGTIRKOTFMAMLKLRQVFLFDGYNF--KRONCEITAL 264
DB 510 rpdeqlasllscdqllkeqsvtemcmrnllqqlknqvlllddykelslbp---vlgkl 566

QY 265 IKENHREKMNVLVTTFTECLRHIOFGALTAEVGDMTEDSAQALIREVLKELA--EGLL 322
DB 567 lqknhlstrcllavrtnraridrryletlllelgaflfynlvcilrlfshnmrltkfm 626
QY 323 LQIQKSRCLRNLMKTPLEFVITCAIOMGESEFHSHTQTFLEFTEYDLIOKNKHKGVA 382
DB 627 vyfngnglqkqkptlvaalcawfygprfddvaavfksymertslrnk-----a 680
QY 383 ASDFTIR-SLDHCGDIALBGVFSHKFDFELQDV--SSVNEVDYLLTGGLCKYTAQREPKY 439
DB 681 laelkatvsscgejalxgfscfctefnddlaeagvdedltnclmskfaagrlrpfy 740
QY 440 KFEHKSQREYTAGRSLSSLLTSHPEEYTKNGYIQKAVSISDITSTSSLLRTGSSV 499
DB 741 rllspaqlcpqayfsmvsehllvllalktagysnt-vaacsprvllqfllgrrlllgaln-q 901
QY 500 EATRA---VMKHLAAVYOHGCLLSIAKRPLMRQESLOSQVKN-----TTEOEI 545
DB 799 pstkagprkivshllhvdn-----keslenlenddyllkhpelslqmqll 843
QY 546 LKAI---NINSEVECGIHL-----YQESTSKALSQEFEEFQGSILYINSGNIPD 593
DB 844 lrglwqicpqayfsmvsehllvllalktagysnt-vaacsprvllqfllgrrlllgaln-q 901
QY 594 YLFD-----FFEHLPNC--ASALDFIKLDFYGG--AMASME 625
DB 902 yfddhpeslslrslhlsirgnktsprahfsvletctfcdksqgprtdqdyasaafepmewe 961
QY 626 -----KAAEDTGINMEBAPERY-IPSRRAVSLFFNMKQEFRLLEVTL 666
DB 962 rnlækednvksyndmgrraapdistgylspkylp-----cleadv 1006
QY 667 RDESKLNQDITTYIGKIRTSATSRLQIKRCAGVAGSLVLSTOCK-NIYSLMWEASPLT 725
DB 1007 ndidvgqdamlellmtvfaeqrlelhlnhsrgflsrlpalelskaavtcsislcls 1066
QY 726 IEDERHITSVNTLKTLSH---DLQNGRLPGGLDLSGLNKLNTKLINDNKKM----- 775
DB 1067 aaegeilltlpsleslevsgtlsgdqqlfpr--ldkflcllelsvdlegnlnvsvipee 1124
QY 776 -----NEEDAIKLAEGLNKLKMKCLEFHL--THLSDIGEGMDYIVKSLSE 818
DB 1125 fprfhmeklllqdsaeqdpkylkvlinspnlnvfhkcnffidsglmtmlys----- 1179
QY 819 PC-DLEETQLVSCCISANAVKYLQNLHNLVKLSTIDE--SENYLEKQDNEALHELIDRMN 876
DB 1180 -ckrlteikfsdfff--qavpfva-slpnfisikllnlegqqfpedeetsekfayllgsls 1235
QY 877 VLEQETLALMLMCGDYGSSLSLKLHLEEVPOVLKLGAKNRLDTERILIGAFGKNPL 936
DB 1236 nlee---lllptgdqlyrvaklllqqcqlnclrlvlsffk-llnddsvelakvaanbgf 1291
QY 937 KNFQDLNLAGN-RVSSDGLWLFMGVFNLLKQLVFFDES---TKFELDPALVRKLSQVLS 992
DB 1292 qklenklsinhkltteeyrnffqaldmmpnlglgdlsrhttecklqagatvkslscvll 1351
QY 993 KTLFLQEARLVKQWOPDDDLSTVI 1015
DB 1352 rlprrlirnmiswlladdaiall 1374

RESULT 10
ID AAY14080 standard; protein; 1295 AA.
XX
XX AAY14080;
AC
XX
XX 20-JUL-1999 (first entry)
XX
XX Gonadotropic hormone protein sequence.
XX
XX Gonadotropic hormone; excessive ovulation animal; transgenic animal;
KW

KW totipotent cell; somatic cell chromosome.
 XX Homo sapiens.
 OS JP11113444-A.
 PN 27-APR-1999.
 XX PD
 XX PF 14-OCT-1997; 97JP-0280830.
 XX PR 14-OCT-1997; 97JP-0280830.
 XX PA (KAGA-) KAGAKU GIUTSU SHINKO JIGYODAN.
 XX (SAKA/) SAKAI H.
 DR WPI: 1999-320709/27.
 XX N-PSDB; AAX58001.
 PT An excessive ovulation animal - useful for improving the
 XX productivity of animals
 XX PS
 Claim 3; Page 14-18; 18pp; Japanese.
 CC This sequence represents a gonadotropic hormone.
 CC The invention relates to an excessive ovulation animal, which is a
 CC transgenic animal with a totipotent cell containing a DNA fragment
 CC containing a promoter sequence and a gonadotropic hormone coding
 CC sequence. The DNA fragment is in the somatic cell chromosome. The
 CC excessive ovulation animal is useful for improving the productivity of
 CC animals. The method can improve the productivity of a useful animal.
 XX
 XX Sequence 1295 AA;
 SQ
 Query Match 11.2%; Score 591.5; DB 20; Length 1295;
 Best Local Similarity 24.4%; Pred. No. 1.3e-40;
 Matches 237; Conservative 167; Mismatches 346; Indels 223; Gaps 38;
 QY 69 FLKSLKEMNYPLFQDLNGQ---SLFHQTSGLDLD----- 100
 DB 344 flgmks-saevrpdlsrgelcelletsesnledsavglvpmaggeaqvfgakn 402
 QY 101 LAODLKLITPPLNF-----YPLGDDIDITNLSKTFEPVLMKRDQHNHR 148
 DB 403 lneqlraaytsafrlmsliddsdladhlqcdlsi---askhlskpv----- 449
 QY 149 VEOTLNGLOALOSPCTIEGESGKSTLQRIAMLMGSGKALKKFFVFLRLS-- 206
 DB 450 qeplvrevgnlnsvmcvegeagsgkvtllkkaflwasgcepllnrfqvlvylslst 509
 QY 207 RAOGGLEFETLCDLDDIGTIRKOTFMAMDLKRVLPFLDGYNEF--RPONCPETIAL 264
 DB 510 rpdqglaslcdqllekgesvtemcmrnllqklnqvlfllddykcelcsipg--vlgkl 566
 QY 265 IKENHRKKNMVIYTTTECLRHIFOGALTAEGVDMTEDSNAQALIREVLKELA--EGLL 322
 DB 567 lqknhlsrtclllavtnrardlrytlellekafrfylvclvclrkllfshnmrlrkfkm 626
 QY 323 LOIOKSRCLRMKTPFVYVITCAIQGSEPHSHQTTLFHPFYDLLQKNKHKGVA 382
 DB 627 vyfgkngslqkltcpflvaicaiahfgypfdrfdvavafksymelstlnk-----a 680
 QY 383 ASDPFR-SLDHCGDLEAGVFNKFDPELDV--SSVNEVDVLLTGLCKYTAQRFKPKY 439
 DB 681 taelkatsvscgelalkgfscfendddlaeaygededlmcclmskftaqlrlpfy 740
 QY 440 KFFHKSQETTAGRRSLSLTSHEPEVTGNGTLOKMWISDITSYSSLLRYTCSSSV 499
 DB 741 rfpafqeflaqmrllellsdqgehdqlylhlkqinspmatvsaynflnyv--ssl 798
 QY 500 EATRA---VMKHLAAYVOHGCLGLSLAKRPLRMROESLOSXKN-----TTEQEI 545
 DB 799 pstcaqpkvlshlhlvdn-----keslensenddylkhpqelslqmql 843

QY 546 LKAI-----NINSEVECGIHL-----YOESTSKSALSOEFAFPOGKSLYINSNIPD 593
 DB 844 lrglwlqcpqayfsmvsehllvalaktaygnt-vaacpffvqlfgtrtlllgalnlg 901
 QY 594 YLFD-----FFEHLPNC--ASALDFIKLDYGG--AAASWE 625
 DB 902 yffdhpeslslrlshfprngktsprahsfvletcdksqyplldqdyasafemewe 961
 QY 626 -----KAEEDTGIIHMEAPETY-IPRAVSLFPNMKQEFRTLEVTL 666
 DB 962 rnlakeednvksymdqgraspalslsgywklspxykfp-----cleavv 1006
 QY 667 RDESKLNKODITYLGIKIFSSATSLRLQIKRCAGVAGSLSVLSTCK-NIYSLWEASPLT 725
 DB 1007 ndldvvgqdmlellmtevfssqfleihsrgfleslpralelskavtkslskls 1066
 QY 726 IEDERHITSTYNKLTISI--DLQNRPLPGGLTDSGNKNTKILMINKI----- 775
 DB 1067 aaegeilltlpsleslevsgtlsgdqifpn--ldkflclkelvdegnlnvsvlpee 1124
 QY 776 -----NEEDAIKLAEGKLNKKMCLFHL--THLSDIGEGMDYIVKSLSSP 818
 DB 1125 fmfhmeklllgiseaypskvlkqnsplnhvthlknctfsdgsimtmivs----- 1179
 QY 819 PC-DLEBIQLVSCLSANAYKILQNLMLVKLSIDL-SENYLEKDGNEALHELIDRN 876
 DB 1180 -cklteikfsdsff--qavpfva-slpntfislkllnlegqgfdeetsekfayllgsls 1235
 QY 877 VLEQLRALMLPMG 889
 DB 1236 nlee--lllptg 1245
 RESULT 11
 ID AAY09540
 ID AAY09540 standard; Protein; 1295 AA.
 XX
 XX AAY09540;
 AC
 XX
 DT 20-JUL-1999 (first entry)
 XX
 DE Human apoptosis inhibiting protein #2.
 XX
 KW Human: apoptosis inhibitory protein; apoptotic disease; diagnosis;
 KW spinal muscular atrophy.
 XX
 OS Homo sapiens.
 XX
 PN JP11116599-A.
 PD 27-APR-1999.
 XX
 PD 14-OCT-1997; 97JP-0280831.
 XX
 PR 14-OCT-1997; 97JP-0280831.
 XX
 PA (KAGA-) KAGAKU GIUTSU SHINKO JIGYODAN.
 XX
 DR WPI: 1999-323531/27.
 DR N-PSDB; AAX56273.
 XX
 XX
 PT New apoptosis inhibitory protein - useful for determining mechanism
 PT of various apoptotic diseases e.g. human spinal muscular atrophy
 XX
 PS Claim 1; Page 8-11; 16pp; Japanese.
 CC The present sequence represents a human apoptosis inhibitory protein.
 CC The apoptosis inhibitory protein is useful for the elucidation of
 CC the mechanism of various apoptotic diseases such as human spinal
 CC muscular atrophy and the diagnosis, the prevention and the treatment
 CC of such diseases.

SQ Sequence 1295 AA;

Query Match 11.2%; Score 591.5; DB 20; Length 1295;
 Best Local Similarity 24.4%; Pred. No. 1.3e-40;
 Matches 237; Conservative 167; Mismatches 346; Indels 223; Gaps 38;

QY 69 FLKSLKEMNYPFDLNGQ-----SLFHQTSGLDLD----- 100
 Db 344 flgmks-saevtplgargelcelletsesnledsIavgpivpemaqgeawfgeakn 402
 QY 101 LAQDLKDLVHPSPFLNF-----YPLGEDIDIFNLKSTFTEPVLKRDQHHR 148
 Db 403 lneqtraaytasaftrmsllldtsdlatclhlgcdsl---askhlsprv----- 449
 QY 149 VEOLTLNGLIALASPCIEGSGKSTLQRIAMMGSGCKAKLTFKVEFLRLS-- 206
 Db 450 qeplvlpvfgnlsvmcvegaagsgktvllkklafIwasgccpIlntfqlvflsIsst 509
 QY 207 RAQGLFETLDCDLIDIGTIRKOTFEMMLKLRORVLFLLDGYNEF--KPNCPETEL 264
 Db 510 rpdgelaIicdqlleksgsvtemcmnllqllknqvlfllddykeicsipq---vlgkl 566
 QY 265 IKENHFRKNMYLVTTTECLRHIRQFALTAEGDMEDSAQALIREVLIKELA---EGL 322
 Db 567 lqknhlstrclliavrtaradirylelleikaflpynvcilrklfshmtlrkfm 626
 QY 323 LQIQSRCLRLNMTPLFVVTCAIQMGSEPHSHOTTTLTFHFVDLLOKKNHKGVA 382
 Db 627 vyfgnqslqkqkprlplvaalcahwfgyrpfidvafkysmerlslnk-----a 680
 QY 383 ASDFR-SLUDHCGDLALGEGVESHKDFELOY--SSVNEDEVLTGLLCKYARQRPKY 439
 Db 681 taelIkavssogelalvgffscfeinddaaegvdedelltnclnskfagllrfy 740
 QY 440 KEFHKSFOEYTAGRRLSLTSHEPEVYKNGYLQKNVSIISDITSTYSLRYTCGSSV 499
 Db 741 rflspafgefIagmrlllelIdsdqehqdlglyhllkqlnspmtvtsaynnflny--sSl 798
 QY 500 EATRA---VMKHLAAYVOHCGCLLGUSTAKRPLMROESLOSQKN-----TTEDEI 545
 Db 799 pstkagpkIvshllhIvhn-----keslenIsenddyllkhpelstlqmgll 843
 QY 546 LKAI-----NINSEVCGIHL-----YQSTSKSALSOEFAFPQKSLYINGNIPD 593
 Db 844 lrglqieqpqayfmsvsehllvllakIaygsnt-vaacsprvlqflgrlcllgalnI-q 901
 QY 594 YLFD-----FPEHLPGNC--ASALDFIKLDFYGG--AMASME 625
 Db 902 yfcdhpesllsrshfprlgnktprahfsvlelctfdksqyrltdqdyasaIepmewe 961
 QY 626 -----KAEDTGIGIMEEPERY-IPSRVSLFPMWKOEFRLLEYTL 666
 Db 962 nlaekednvksymdqrraapdsIstgywkIspkgykIp-----clelv 1006
 QY 667 RDFSRLNRDITVYLGKIFSSATSLRLQIRCAAGVAGSLVLTSTCK-NIYSIMVASPLT 725
 Db 1007 ndldvvgdmeIImbvtvsaagrlhlnhargfleslrpaleIskaavtcsIsklleIs 1066
 QY 726 IEDERHTSVTNLKTLSLH--DLONQRLPGGLDLSLGNLKNLTKLINDNIM----- 775
 Db 1067 aaegallltlpsleslevsgltsgdqqlfpr--ldkflclkelavsdlegnlnvIsIpee 1124
 QY 776 -----NEEDAIKLAEGSLKNLKKMCLFHL--THLSIDIGGMQYIVASLSSE 818
 Db 1125 fpuflhmeklllqIsaeydpsklvllqnsplhvfhlkcnffsdqslmImvls----- 1179
 QY 819 PC-DLEETIOLVSCSAANAVKILAOHLNHLVLTSLDL-SENYELDKDNEALHELDIMN 876
 Db 1180 -ckktelkfsdIsff--qavpIva-sIpnfIsIkllnlegqgfdeecsekIayllgsls 1235
 QY 877 VLEQTLTALMLPWG 889
 11: 1:1:1 1

Db 1236 mlee---lllptg 1245

RESULT 12

ID AAR98217 standard; Protein; 1232 AA.

AC AAR98217;

DT 30-DEC-1996 (first entry)

DE Neuronal apoptosis inhibiting protein.

KW Neuronal apoptosis inhibiting protein; human; NAIP; chromosome 5q13; YAC;
 KW yeast artificial chromosome; spinal muscular atrophy; mammalian cell;
 KW autosomal recessive; neurodegenerative disorder; alpha motor neuron; SMA;
 KW spinal cord; proximal voluntary muscle; therapy; apoptotic mechanism.

OS Homo sapiens.

BN W09612016-A1.

PD 25-APR-1996.

PF 17-OCT-1995; 95WO-CA00581.

PR 19-DEC-1994; 94CA-2138425.

PR 18-OCT-1994; 94GB-0021019.

PA (SHKJ) RES DEV CORP JAPAN.

PA (UYOT-) UNIV OTTAWA.

PI Ikeda J, Korneluk RG, Mackenzie AE, Mahadevan MS;

PI McLean M, Roy N.

DR WPI: 1996-222003/22.

NR N-PSDB; AAT30092.

PT Neuronal apoptosis inhibitor protein gene - used to develop prods.

PT for use in the diagnosis and therapy of spinal muscular atrophy

PS Claim 3; Page 68-70; 113pp; English.

XX This sequence represents the human neuronal apoptosis inhibitor protein
 CC (NAIP). The cDNA encoding this sequence was found on a region of the
 CC human chromosome 5q13. This sequence was isolated from a yeast
 CC artificial chromosome (YAC) contig containing the D5S435-D5S112 interval
 CC of the chromosome 5q13. Mutations in the NAIP gene, are causative of
 CC spinal muscular atrophy (SMA) types I, II, and III. SMAs are a group of
 CC autosomal recessive, neurodegenerative disorders. SMAs are classified
 CC into the three types based upon the age of onset (with type I being the
 CC severest form with the earliest age of onset). All three types are
 CC characterised by the degeneration of the alpha motor neurons of the
 CC spinal cord manifesting as weakness and wasting of the proximal voluntary
 CC muscles. The most common mutations of the NAIP gene sequence are thought
 CC to be deletions of exons 5 and 6, and reductions in the copy number of
 CC the gene. The NAIP gene, (and primers and probes based on it) can be
 CC used for the diagnosis of SMA, and for directing the formulation of
 CC conventional and genetic therapies for SMA. Identification of genes
 CC showing homology with the NAIP locus, and proteins that interact with
 CC NAIP can be used in the elucidation of apoptotic mechanisms in mammalian
 CC cells.

SQ Sequence 1232 AA;

Query Match 10.6%; Score 561.5; DB 17; Length 1232;
 Best Local Similarity 24.1%; Pred. No. 4.1e-38;
 Matches 226; Conservative 155; Mismatches 341; Indels 215; Gaps 33;

QY 69 FLKSLKEMNYPFDLNGQ-----SLFHQTSGLDLD----- 100
 Db 344 flgmks-saevtplgargelcelletsesnledsIavgpivpemaqgeawfgeakn 402
 11: 1:1:1 1

XX 09-AUG-2001.
 PD 30-JAN-2001; 2001WO-US00663.
 PE
 XX 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 WPI; 2001-488897/53.
 XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human placenta -
 PS Claim 27; SEQ ID No 30571; 654pp; English.
 CC The present invention relates to single exon nucleic acid probes (SENP;
 CC see AI13315-AI57546). The present sequence is a peptide encoded by one
 CC such probe. The probes are useful for producing a microarray for
 CC predicting, measuring and displaying gene expression in samples derived
 CC from human placenta. The probes are useful for antenatal diagnosis of
 CC human genetic disorders.
 XX
 SQ Sequence 73 AA;

Query Match 7.0%; Score 372; DB 22; Length 73;
 Best Local Similarity 100.0%; Pred. No. 5.2e-24;
 Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 944 LAGNRVSSDCGLAFMGVFENIKLVFPDFSTKEFLPPALVRKLSGYLSKTLFLOEARLY 1003
 Db 1 lagnrvasdgylafmgvfenkqlvtfdfstckelipopalvrklsqylskltflgearlv 60

OY 1004 GWFPPDDDLVSIT 1016
 Db 61 gwqfdddlsvlt 73

RESULT 15

AAM05439
 ID AAM05439 standard; Protein; 73 AA.
 XX
 AC AAM05439;

DT 09-OCT-2001 (first entry)

DE Peptide #4121 encoded by probe for measuring breast gene expression.

XX Probe: human; breast disease; breast cancer; development disorder;
 KM inflammatory disease; proliferative breast disease; non-carcinoma tumour.
 XX
 OS Homo sapiens.

PN WO200157270-A2.

XX 09-AUG-2001.

PD 29-JAN-2001; 2001WO-US00661.

PE 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 WPI; 2001-476286/51.
 XX Novel single exon nucleic acid probe used to measuring gene expression
 PT in a human breast -
 PS Claim 27; SEQ ID No 14179; 322pp; English.
 CC The present invention relates to novel single exon nucleic acid probes
 CC (see AI100010-AI10067). The present sequence is a peptide encoded by one
 CC such probe. The probes are useful for measuring human gene expression in
 CC a human breast sample, where the probe hybridises at high stringency to a
 CC nucleic acid expressed in the human breast. The probes are useful for
 CC predicting, diagnosing, grading, staging, monitoring and prognosing
 CC diseases of the human breast, particularly those diseases with polygenic
 CC aetiology. The diseases include: breast cancer, disorders of development,
 CC inflammatory diseases of the breast, fibrocystic changes, proliferative
 CC breast disease and non-carcinoma tumours.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 73 AA;

Query Match 7.0%; Score 372; DB 22; Length 73;
 Best Local Similarity 100.0%; Pred. No. 5.2e-24;
 Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 944 LAGNRVSSDCGLAFMGVFENIKLVFPDFSTKEFLPPALVRKLSGYLSKTLFLOEARLY 1003
 Db 1 lagnrvasdgylafmgvfenkqlvtfdfstckelipopalvrklsqylskltflgearlv 60

OY 1004 GWFPPDDDLVSIT 1016
 Db 61 gwqfdddlsvlt 73

Search completed: March 25, 2002, 10:56:18
 Job time: 78 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 25, 2002, 10:56:20 ; Search time 14.93 Seconds
(without alignments)
2514.721 Million cell updates/sec

Title: US-09-697-089-2

Perfect score: 5281

Sequence: 1 MNFKDNRSLRQRMGMVY.....MQPDDDLVITGAFKLVTA 1024

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 3664827 residues

Total number of hits satisfying chosen parameters: 100055

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match	Length	ID	Description
1	665.5	12.6	1403	1	BIRF_MOUSE	Q911b6 mus musculu
2	657	12.4	1402	1	BIRG_MOUSE	Q911b3 mus musculu
3	653	12.4	1403	1	BIRI_HUMAN	Q13075 homo sapien
4	637.5	12.1	1403	1	BIRE_MOUSE	Q91016 mus musculu
5	637.5	12.0	1447	1	BIRB_MOUSE	Q99u4 mus musculu
6	631.5	12.0	1403	1	BIRA_MOUSE	Q99u43 mus musculu
7	180.5	3.4	1078	1	C2TA_MOUSE	P79621 mus musculu
8	176	3.3	862	1	AACT_DICDI	P05095 dictyosteli
9	168	3.2	3911	1	AKA9_HUMAN	Q99996 h a kinase
10	164	3.1	1060	1	EG51_XENLA	P28025 xenopus lae
11	152.5	2.9	2418	1	SPCA_HUMAN	P02549 homo sapien
12	150	2.8	1277	1	YMW6_YEAST	Q04264 saccharomyc
13	149.5	2.8	3210	1	CENF_HUMAN	P49454 homo sapien
14	148	2.8	2167	1	BEM2_YEAST	P39660 saccharomyc
15	146.5	2.8	2493	1	YBAA_YEAST	P35194 saccharomyc
16	146	2.8	1682	1	MSPI_PLA3	P19598 plasmodium
17	145.5	2.8	4196	1	DYHC_SCHRO	O13290 schizosacch
18	143.5	2.7	1315	1	CHAO_DROME	P12024 drosophila
19	143.5	2.7	2869	1	RBP1_PLA3	Q00798 plasmodium
20	143	2.7	875	1	ZIP1_YEAST	P31111 saccharomyc
21	141.5	2.7	1130	1	C2TA_HUMAN	P33076 homo sapien
22	141	2.7	2477	1	SPCN_CHICK	P07751 gallus gall
23	140.5	2.7	326	1	YMN29_CAEEL	P45969 caenorhabdi
24	140	2.7	1875	1	MPLI_YEAST	Q02455 saccharomyc
25	139.5	2.6	567	1	YE28_MERVA	Q38823 methanococc
26	139.5	2.6	5430	1	ACF7_HUMAN	Q99u43 homo sapien
27	138.5	2.6	1526	1	MYSC_SCHRO	Q99u43 schizosacch
28	138.5	2.6	4540	1	DYHC_PARPE	Q27171 paramacium
29	138	2.6	2710	1	TOXA_CLODI	P16154 clostridium
30	138	2.6	5327	1	ACF7_MOUSE	Q99u43 mus musculu
31	137	2.6	1256	1	FULI_DROME	Q24020 drosophila
32	136	2.6	999	1	RUK5_ARATH	P47735 arabidopsis
33	136	2.6	1048	1	HPM1_YEAST	P51979 saccharomyc

34	135.5	2.6	3321	1	KEND_HUMAN	O95613 homo sapien
35	135	2.6	2230	1	GOG4_HUMAN	O13439 homo sapien
36	134.5	2.5	1928	1	MYSL_YEAST	P08964 saccharomyc
37	134	2.5	567	1	GPV_RAT	O08770 rattus norv
38	134	2.5	959	1	VDP_RAT	P41542 rattus norv
39	133.5	2.5	456	1	RINI_RAT	P29315 rattus norv
40	133.5	2.5	1818	1	Z294_HUMAN	O94822 homo sapien
41	133	2.5	1325	1	G160_MOUSE	P55937 mus musculu
42	133	2.5	2238	1	RRLP_BUNTW	P20470 bunyamwera
43	132.5	2.5	582	1	SHO2_MOUSE	O88520 mus musculu
44	132.5	2.5	1057	1	EG5_HUMAN	P52732 homo sapien
45	132.5	2.5	1648	1	Y39H_YEAST	P47171 saccharomyc

ALIGNMENTS

RESULT 1

ID BIRF_MOUSE STANDARD: PRT: 1403 AA.

AC Q911b6: P81704; 009122; 009121;

DT 20-AUG-2001 (Rel. 40, Created)

DT 20-AUG-2001 (Rel. 40, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE BACULOVIRAL IAP REPEAT-CONTAINING PROTEIN 1F (NEURONAL APOPTOSIS INHIBITORY PROTEIN 6)

CN BIRCLF OR NAIP6 OR NAIP-RS4.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI_TaxID=10090;

[1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20414747; PubMed=10958627;

RA Endrizzi M.G., Hadinoto V., Growney J.D., Miller W., Dietrich W.F.;

RT "Genomic sequence analysis of the mouse Naip gene array.";

RL Genome Res. 10:1095-1102(2000).

RN [2]

RP SEQUENCE OF 82-168 FROM N.A.

RC STRAIN=129/SVJ;

RX MEDLINE=97131520; PubMed=8975718;

RA Scharf J.M., Damron D., Frisella A., Bruno S., Beggs A.H., Kunkel L.M., Dietrich W.F.;

RT "The mouse region syntenic for human spinal muscular atrophy lies within the 1q91 critical interval and contains multiple copies of Naip exon 5.";

RL Genomics 38:405-417(1996).

CC -1- FUNCTION: PREVENTS MOTOR-NEURON APOPTOSIS INDUCED BY A VARIETY OF SIGNALS.

CC -1- SIMILARITY: CONTAINS 3 BIR REPEATS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AF242431; AAF82751.1; -

DR EMBL; U66327; AAC52975.1; -

DR MGD; MGI:1298222; Birc1f.

DR InterPro; IPR001370; BIR.

DR Pfam; PF00653; BIR; 3.

DR SMART; SM00238; BIR; 3.

DR PROSITE; PS01282; BIR_REPEAT_1; 2.

DR PROSITE; PS0143; BIR_REPEAT_2; 3.

KW Apoptosis; Repeat; Multigene family.

FT REPEAT 60 127 BIR 1.

FT REPEAT 159 227 BIR 2.

FT REPEAT 278 345 BIR 3.

SQ SEQUENCE 1403 AA: 159823 MW; 9DA912503358C4E9 CRC64;

Query Match	12.6%;	Score 665.5;	DB 1;	Length 1403;
Best Local Similarity	24.1%;	Pred. No. 1.3e-30;		
Matches 250;	Conservative 186;	Mismatches 414;	Indels 187;	Gaps 37;

QY	101	LAODKRLUTYPPSLNF-----YPLGEDDILJFNFKSTPEBVLMRKQDHNH	148
Db	403	LSBDRRTYKTSCHMHNLFVCCSSLSCTDHLLOCVSI-----SKHVSOPV-----	449
QY	149	VEOLTLNGLLOALQSPCIIIEGSGKGRKSPYLORIAMJMGSGKCALTKREYFVLRLS--	206
Db	450	OGALTFIEVENSINLSWVCYGEAGSGKTFELKIAFLMASCPCPLRYRQOLVYLSISI	509
QY	207	RAQSGFETLQDOLDLPGIRIKROTAMUKIKRQGVLELLDQYMEFK--PONCEIDAL	264
Db	510	TPDQGLANITCTOLLLGAGGCISEVCISSTIQOLQOHQVLELLDYGSLASLPQ--ALHFL	566
QY	265	IKENHRKKNVATYTTTECRLRHITROFGALTYAEGDVTEDSAQALIREVLIKEI--AEGIL	322
Db	567	ITKNVLEFRTCLLIAVHNRRDIRPLTGLSLEIOEPFVNTYFVLKRPFSSHIIQVCEKLI	626
QY	323	LOIQSKRLNMLKPTLEFVITC-AIQMBSEPHSTQTLTFHTEYDLLIOKNHKHKGV	381
Db	627	IFYSEENDLOGVYVTPLEFVAAYCNDMNONASADDDQVDTLEHSMOYLSLKYK-----A	681
QY	382	AASFIRSLDHCGLALEEGVSHKHPPELODY--SSVNDVLTLLTGLCKCTAQRKFRKY	439
Db	682	TAESLOATVSSCGOALATGLFSSCFEPNSDDIAEACVDYVAKLITFLMSKTFQNRKYV	741
QY	440	KPFHKSFOEYTAGRRLSLTLTSHEPEVTKGNGYLOKMYISIDITYSSTSLRYCYG--SS	498
Db	742	RFLGPLRQEFLLAARVRELTLLSDROEQDQGLGYLRLQDSPKAIANSFNIFLYVSSHSS	801
QY	499	VEATRAVYKHLAANYONGCLLGISIAKRLPMROESIQSKNTBOE-----	544
Db	802	SKAAPTYVSHL-----LOLVDEKSELENNSENEDEYKALHPOTFLMQ	843
QY	545	-----ILKAININSFEGCIIHLYQ-----ESTSKSALSOEFAPFGKSLYNSGNI	591
Db	844	FVRLGMLVSPESSEFY--SEHLLRLALIFAYSNSNVACSPFILOPLRGRTALARVNL	901
QY	592	PQYLEDFEPEHLPCNCASALDFIKLDFYGAGAMWEKAEDTGGIIMHEAPETYIPSAVLS	651
Db	902	-EY---FEMDH-PESILLRLSLKVSINGNMKSSYDVSEFKT--YEENLOPRAINEEYSA	953
QY	652	F---FNNKOER-----TLETYLRPFKSLN	673
Db	954	FHVHSEKRRNFADODEITIKYENIMWRPALPDISEGYWNLSPKPCIKPLEYOVANNMGPAD	1013
QY	674	KQDITVYLGKIFESSATSLRLQIKRCAGVAGSLSVLSTCK-NIYSLMVEASPLTIDERBHI	732
Db	1014	QALQVIMEVFSASQSTIEFHLFNSSGFLSIRPALELSKASVTKCSMSRLSLSRAEQEL	1073
QY	733	TSVYNLKLST-----HDLON-----ORLPG-----LTDLSGINKNLTK	767
Db	1074	LTLPALQSLVESETNOLPDQLFHNHLKFLGLKELCVRLDGKPDVLSVPEEFLNHNHEK	1133
QY	768	LIMONIKMNEEDATKLAIEGLKNTKKMCFLPHLTSIQEGMOTYIKSSISEPC-DLEIO	826
Db	1134	LSIT--STESDLSKLKVFQONPPNLAHVHLK--CDFLSNCSLMTALAS--CKKLREIE	1187
QY	827	LVSCCLSA-NAVKILAQNLHNLVKLSILD--SENYLEKDGNEALHELIDRNVLEQOLAL	884
Db	1188	FSGGCFEAMFVNILP---NFVSLKLSLKGQGFADKRTSKPQALGSLUNLEE--L	1240
QY	885	MLPMGCDVQGSLSLKLHLEBPOLYKGLKMKRLTDTETIRLIGAFEGKNPLKFNQOLNL	944
Db	1241	LVPTGDDJHOYAKLIVQOCLOLPCRLVLAFFH--ILDESEVIEIGEAATSGSFQKLENDI	1299
QY	945	AGN-RVSSDGLMAVGEFENIKOLVFFDFSTKEFLP-----DPALYRKLSQYLSLFLQ	998
Db	1300	SMNHKITEEGYRNFQALDNLPLNLOHNLICRN--IGRQIQOVATTVYKALGHCYVSLPSLT	1357

```
QY 999 EARLVGWFDDDDLSVI 1015
    :: | |:: | |
Db 1358 RLGMLSWLLDEEDMKVI 1374
```

	RESULT	2
BIRG_MOUSE	BIRG_MOUSE	
ID	BIRG_MOUSE	STANDARD; PRT: 1402 AA.
AC	O9JIB3;	
DT	20-AUG-2001 (Rel. 40, Created)	
DT	20-AUG-2001 (Rel. 40, Last sequence update)	
DT	20-AUG-2001 (Rel. 40, Last annotation update)	
DE	BACULOVIRAL IAP REPEAT-CONTAINING PROTEIN 1G (NEURONAL APOPTOSIS INHIBITOR PROTEIN 7).	
GN	BIRCIG OR NAIP7.	
OS	Mus musculus (Mouse).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
OX	NCBI_TaxID=10090;	
RP	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=20414747; PubMed=10958627;	
RA	Endrizzi M.G., Hadnoto V., Gromey J.D., Miller W., Dietrich W.F.:	
RT	"Genomic sequence analysis of the mouse Naip gene array."	
RL	Genome Res. 10:1095-1102(2000).	
CC	-I- FUNCTION: PREVENTS MOTOR-NEURON APOPTOSIS INDUCED BY A VARIETY OF SIGNALS.	
CC	-I- SIMILARITY: CONTAINS 3 BIR REPEATS.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - between the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).	
CC	-----	
DR	EMBL: AF242433; AAF82749.1; -	
DR	MGB: MGI:1858256; Birc1g.	
DR	InterPro: IPR001570; BIR.	
DR	Pfam: PF00653; BIR; 3.	
DR	SMART: SMO0238; BIR; 3.	
DR	PROSITE: PS01282; BIR_REPEAT_1; 2.	
DR	PROSITE: PS0143; BIR_REPEAT_2; 3.	
KW	Apoptosis; Repeat; Multigene family.	
FT	REPEAT 60 127 BIR 1.	
FT	REPEAT 159 227 BIR 1.	
FT	REPEAT 278 345 BIR 2.	
FT	REPEAT 278 345 BIR 3.	
SO	SEQUENCE 1402 AA; 159662 MW; CIDEFBA359893E0D CRC64;	

Query Match 12.4%; Score 657; DB 1; Length 1402;
Best Local Similarity 24.3%; Pred. No. 4e-30;
Matches 253; Conservative 184; Mismatches 409; Indels 194; Gaps 38

OY	101	LAODLKDXYHRSFNF-----XJESGDIDIFNFKSTPERLYMKRDONNR	148
	403	LSBOARDRYTSTSCNMHLPVSSSLSTGDNHSLCDVSTI---SKNIGRV-----	449
OY	149	VEOLTLNGLALDOSPCTIBESBKGSTLDRJAMLMGSGKALKTKFVEFLRLS--	206
Db	450	OGALTRPEVFNLSVYMCVEEBSAGKTEFLKRAFAFMASGCPRLRFOLVYLSSTI	509
OY	207	RAOGSLFETLDLDLDYRGTRKOTPMALIKLYRORVLEFLDGCYNEPK--POMSREIAL	264
Db	510	TPROGDLDNICTOLLAGGCSSEVCLASSIOTQLOHVLFLDDOVSGLASLRFQ--ALNTL	566
OY	265	IKENHREFKNMAYIVTTTPECLNHPQFALTRYVDQDMEDSDQALREVILKEL--AGLL	322
Db	567	ITKNLEFTCLLDLKNVTRVDIPRYLQTSLEIDEPFRPVFLKRFESHDIIICVKLI	628
OY	323	LOIQKSRCLRMIMKTRPLEFVVTIC--AIOMGESEFHSHTQPTLLEHTFYDDLIOKKNKHNGV	381

Db 627 IYSENKDLQYKTPLEVAANCNDNMNANASQDDVTLFHSYQVLSIKYK-----A 681
 QY 382 AASDFIRSLDHGDLALBEGVFSHKPFELQDY--SSVNEDVLLTGLCKYTAQREPKY 439
 Db 682 TAEISQATVSSGCGLATGLFSSCFEFNDDLAEGVDDVLTFLMSKFLAQMLRPY 741
 QY 440 KFFHKSFOETYAGRLSSLSLSHEPEVYTKNGVYLOKMWISIDITSTYSSSLRYTCG--SS 498
 Db 742 RFLGFLFOEFLAVALRLLTESSDROEDDLGLYLRQIDSPKAIANSFNIFLYVSSHSS 801
 QY 499 VVATRAVKKHLAAVYQHCGLLGLSLAKRPLMRQESLQSKNTDE----- 544
 Db 802 SAAFTVSSH-----LQVDEKESLENNSENEYDKLHPQFTLMFQ 843
 QY 545 -----ILKAININSFVECGIHLVQ-----ESTKSALSOEFAFPQSKLYNSGNI 591
 Db 844 FVRGLMTVSPESFSSFV--SEHLRLALIFAESNTVAECSPFIQFLGRRLARVLT 901
 QY 592 PDYLFDFEHLDPNCASALDFIKLDYFGAMASWEKAEDTGIMHEAPETIYPSRAVSL 651
 Db 902 -EY---FMDH--PESILLRLSLKVSINGNMSSYVDYSEFT--YFENLQPPAINEEYTS 953
 QY 652 F---FNMKQEPF-----TLEVTLRPFSKLN 673
 Db 954 FEHVSEMRNRNFODEEITIKNTENINPRALPDISEGYMNLSPKPKIPKLEVOVNMKPAD 1013
 QY 674 KODITYLAKIFFSATSLRLQIKRCAGVAGSLVSTCK-NIYSIMVAPSLTIDERHI 732
 Db 1014 QALQVLMVEFASQSIIEFHLENSGFLSESTRPALBELSKASTKCSMRLELSRAEOLL 1073
 QY 733 TSVNMLKLTST-----HDLYN-----QRLPG-----LTDLSGNLKNLTK 767
 Db 1074 LTLPALQSLSEVETNQLPDOLFHNHKEFLGELKELCVRLDGKDYVLSPEEFLNLHMEK 1133
 QY 768 LIMDNINKNNEPAIKLAGSKMLKMKCLFHLTHSDIGBMGYIKYLSSEFC-DLEEQ 826
 Db 1134 LSIRF--STESLSKLVKFIQMFNPLVHFHLK--CDPLSNCSLMTALAS--CKLRELE 1187
 QY 827 LVSCSLA-NAVKILAONLHNLKLSILD--SENYLEKDGNEALHELIDRMNVLEQTLAL 884
 Db 1188 FSGQGFEMATFNILP-----NPVSLKILSLKQGFADKETSFKFQALGSLRNLDE---L 1240
 QY 885 MLPWCCDVQGISLILKLEVEYPOLVKGKLNKWLDTDEIRILGAFEGKNPLKNFOOL-N 943
 Db 1241 LVPTDGDGHQVAKLIVROCLQPCRLVLAFAHIDDESIEIGATSG----SFQKLEN 1295
 QY 944 L---AGNRVSSGMLAFMGVFEENLQVLFEPDSTKEFLP-----DPALVRLKSLQVLSKLT 995
 Db 1296 LDISNMHKTTEGYNRFQALDNLPLQMLNLCRN--IPGRIOVATTVKALCHCVSRLP 1353
 QY 996 FLOEARLVGMOFDDDLVSI 1015
 Db 1354 SLTRGLGMLLDEEDMKVI 1373

RA Roy N., Mahadevan M.S., Mclean M., Shuler G., Yaraqhi Z.,
 RA Farahani R., Baird S., Besner-Johnston A., Lefebvre C., Kang X.,
 RA Saikh M., Aubry H., Tamai K., Guan X., Ioannou P., Crawford T.O.,
 RA de Jong P.J., Surh L., Ikeda J., Korneluk R.G., Mackenzie A.,
 RT "The gene for neuronal apoptosis inhibitor protein is partially
 RL deleted in individuals with spinal muscular atrophy.";
 RN Cell 80:167-178(1995).
 RP [2]
 RP SEQUENCE FROM N.A., AND REVISIONS.
 RC TISSUE-Brain:
 RX MEDLINE-98163755; PubMed-9503025;
 RA Chen Q., Baird S.D., Mahadevan M., Besner-Johnston A., Farahani R.,
 RA Xuan J.-Y., Kang X., Lefebvre C., Ikeda J.-E., Korneluk R.G.,
 RA Mackenzie A.E.;
 RT "Sequence of a 131-kb region of 5q13.1 containing the spinal muscular
 RT atrophy candidate genes SMN and NAIP.";
 RN Genomics 48:121-127(1998).
 RL [3]
 RP SEQUENCE OF 386-623 FROM N.A.
 RA der Steege G., Draaijers T.G., Grootscholten P.M., Oslinga J.,
 RA Anzevino R., Veldma I., Brahe C., Scheffer H., van Ommen G.J.B.,
 RA Buys C.H.C.M.;
 RL Submitted (MAY-1995) to the EMBL/Genbank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 222-1403 FROM N.A.
 RA Jones K., Graves T., McPherson J.;
 RL Submitted (JUN-1998) to the EMBL/Genbank/DBJ databases.
 RN [5]
 RP FUNCTION.
 RC TISSUE-Liver:
 RX MEDLINE-96149249; PubMed-8552191;
 RA Liston P., Roy N., Tamai K., Lefebvre C., Baird S., Cherton-Horvat G.,
 RA Farahani R., Mclean M., Ikeda J., Mackenzie A., Korneluk R.G.;
 RT "Suppression of apoptosis in mammalian cells by NAIP and a related
 RT family of IAP genes.";
 RL Nature 379:349-353(1996).
 CC -I- FUNCTION: PREVENTS MOTOR-NEURON APOPTOSIS INDUCED BY A VARIETY OF
 CC SIGNALS.
 CC -I- TISSUE SPECIFICITY: EXPRESSED IN MOTOR NEURONS, BUT NOT IN SENSORY
 CC NEURONS. FOUND IN LIVER AND PLACENTA, AND IN A LESSER EXTENT IN
 CC SPINAL CORD.
 CC -I- DISEASE: MUTATED OR DELETED FORMS OF NAIP HAVE BEEN FOUND IN
 CC INDIVIDUALS WITH SPINAL MUSCULAR ATROPHY TYPE I (SMA TYPE I). SMAS
 CC ARE FATAL AUTOSOMAL RECESSIVE DISORDERS SUBCLASSIFIED AS TYPE I
 CC (MERDING-HOPFMANN DISEASE), TYPE II (INTERMEDIATE FORM), AND TYPE
 CC III (WOLFFHART-KUGELBERG-MELANDER DISEASE) BASED UPON THE AGE OF
 CC ONSET AND CLINICAL SEVERITY. THESE NEURODEGENERATIVE DISORDERS ARE
 CC CHARACTERIZED BY DEGENERATION OF LOWER MOTOR NEURONS, LEADING TO
 CC PROGRESSIVE PARALYSIS MUSCULAR ATROPHY. CONCERNS I IN 6000
 CC NEBORN.
 CC -I- SIMILARITY: CONTAINS 3 BIR REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U19251; AAC52045.1; -;
 DR EMBL; U80017; AAC52047.1; -;
 DR EMBL; U21913; AAA64504.1; -;
 DR EMBL; AC005031; AAC62261.1; -;
 DR MIM; 600355; -;
 DR InterPro: IPR001370; BIR.
 DR Pfam; PF00653; BIR; 3.
 DR SMART; SM00238; BIR; 3.
 DR PROSITE; PS01282; BIR_REPEAT_1; 3.
 DR PROSITE; PS0143; BIR_REPEAT_2; 3.
 KW Apoptosis; Repeat. 127
 FT REPEAT 60 127 BIR 1.
 FT REPEAT 159 227 BIR 2.

```

FT REPEAT 278 345 BIR 3.
FT CONFLICT 222 223 PK -> YR (IN REF. 4).
FT CONFLICT 386 387 VP -> ST (IN REF. 3).
FT CONFLICT 535 535 M -> V (IN REF. 3).
FT CONFLICT 553 553 Y -> H (IN REF. 3).
FT CONFLICT 1228 1231 MISSING (IN REF. 4).
SQ SEQUENCE 1403 AA: 159613 MW: 566304C154DA5E64 CRC64;

Query Match 12.4%: Score 653; DB 1; Length 1403;
Best Local Similarity 23.7%; Pred. No. 6.9e-30;
Matches 263; Conservative 197; Mismatches 410; Indels 238; Gaps 42;

QY 69 FLKSLKEMNPLFQDLNGQ---SLFHQTSGLDD-----100
D 344 FLQNMKS-SAEVPPDLQSRGELCELETSESNEIDSAVPIPEMAQGBAQFQAKN 402
QY 101 LAODKLUYTPFSLNF-----YPLGEDIITNKSTFPEVPMKRDQHHR 148
D 403 LNEQLRAAYTSASFRHMSLDDISSDLATDHLGCDLSI---ASKHISKPV-----449
QY 149 VEOTLNGLLAOLQSPCITIGESGKSTLLORIAMLGSGKALKPKFVFLRLS--206
D 450 QEPVLVEVGNLNSVNCVEGASGKTVLLKAFIAMSGCCPLNRFQVLVLSST 509
QY 207 RAOGLEPTELCDLIDPRTKQTFMAMLEKLQRYVFLLDGYNE--KRONCPRIEAL 264
D 510 RPEGLASIIICDQLLEKESVTEKCMENIIQQLKNQVFLLDVYKEICSIPO--VIGKL 566
QY 265 IKENHRKKNVYITTTRECLRHIOFQALRAEVGDMEDSNOALIREVLKELA--EGLL 322
D 567 IOKNHLSTRCLLAVNRNRRADIRYLETLEKAPFPYTVCLRKLFSSHMRKRKF 626
QY 323 LQIQSRCLNLMKTPFVVTTCALOMGESEFHSHTQTTLFHTFYDLLIOKNKHGVA 382
D 627 VYEKNGSLQKIQKTPFVAICAHMFQYPRDPDVAVKSMELSLRNK-----A 680
QY 383 ASDPFR-SLDHCDLAEGVFSHKFDELDV--SSYNEDVLLTGLCKYTAORFKPY 439
D 681 TAEIKATVSCGELAKGFECSCEFNDDLAEGVDEEDLTMCLMSKRTAORLPFY 740
QY 440 KFFKSOEYTAGRRLSLSLTSHPREVTNGNGYLOKMWISDITSYSSSLRTGSSV 499
D 741 RFLSPAQOEFLAGKRLLELSDROEODLGLYHLKQNSPMATVSAVNNFLNTV--SSL 798
QY 500 EATRA---VMKHLAAYVOHOGCLGLSLAKRPLMROESLOSAYKN-----TTEOEI 545
D 799 PSTKAGPKIVSHLHLVDN-----KESLENISENDVYLKHQPEISLQML 843
QY 546 LKAI---NINSVECGIHL-----YOEISTKSALSOEFEPQOKSLYINSQINPD 593
D 844 LRGLMOIQPOAFYSGVSEHLLVLAUKTAYOSNT-VAACSPFVLQFLOGRTLLGALND-Q 901
QY 594 YLFP-----PFHLFNC--ASALDFIKLDFEGG--AMASWE 625
D 902 YEFDPHPSLSLNSIHPPINGNKTSPPRAHFSVLETCEDKSOVPTIDODYSAFEPNME 961
QY 626 -----KAEDTGGIIMEBAPETY-IPSRHVSILFFNMKOEFTLEVTLL 666
D 962 RNLAERKEDNKSVMQDMQRASPDLSGTGWKSPKQYIP-----CLEVDY 1006
QY 667 RDSKLMKODITYIGKFFSATSLRLQIKRCAGVAGSLVLSYCK--NIYSLWASPLT 725
D 1007 NDIIVGQDMLEIMTFVSQSRIELHNLNSRGFIESIRALELSKASVYKCSISKLELS 1066
QY 726 IEDERHITSVNLKTLISH--DIQONRLPGGLDSIGNKLNKLMDNKKM-----775
D 1067 AAEDELLTLPSLESLSGTSQSDQIFPV--LDKFLCKEISLVDEGNWFSVPIPEE 1124
QY 776 -----NEEDAIIKLAEGKLNKKMLCFHL--THLSDIGENDYIYKSLSE 818
D 1125 FPNFHNMEKLLIIQISAEVDPKLVKLLQNSPNLHVPHLKCNFFSDFPSLMMVLVS-----1179

```

```

QY 819 PC-DLEBIQVSCCLSANAVKILAOHLNKLSTIDL-SENYLEKDGNEALHELDRN 876
D 1180 -CKMLTEIKFSDESEF--QAVPEVA-SLPNFIKILNLEEQCPDEBTESEKAYIIGLSL 1235
QY 877 VLEQILALMPPMCDCVOGSLSSLLKHLEVPOLVKLGKMMRLTDR-----ELRIIGAPF 931
D 1236 NLEE---LILPTDGDYIRVAKLLIQQCQOLHCLRVLSFEKTLNDDSVETAKVAISGGF- 1291
QY 932 GKNPLKNFOQLNLAGN-RVSSDGMFLAFMGVFENIKOLVPEDFS---TKRELPPALVRKL 987
D 1292 -----QKLENLKLSTINKKITEBEGYRNFPQALDMPNQLDELISRHTTECKAKATYKSL 1346
QY 988 SQVLSKLTFLQEARLVGQWQFDDDLSTV 1015
D 1347 SQCVLRILRPLIRLNMLSMLDADIALL 1374

RESULT 4
BIRE_MOUSE
ID BIRE_MOUSE STANDARD; PRT; 1403 AA.
AC Q9R016; Q9R029; P81703; 009122; 009121;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE BACULOVIRAL IAP REPEAT-CONTAINING PROTEIN IE (NEURONAL APOPTOSIS
INHIBITORY PROTEIN 5).
GN BIRCIE OR NAIP5 OR NAIP-RS3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
ON NCBI-TaxID=10090;
RX MEDLINE=99431676; PubMed=10501978;
RA Huang S., Scharf J.M., Growney J.D., Endritz M.G., Dietrich W.F.;
RT "The mouse Naip gene cluster on chromosome 13 encodes several distinct
functional transcripts."
RL Mamm. Genome 10:1032-1035(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RX MEDLINE=99417674; PubMed=10486205;
RA Endritz M., Huang S., Scharf J.M., Kelter A.R., Wirth B.,
RA Kunkel L.M., Miller W., Dietrich W.F.;
RT "Comparative sequence analysis of the mouse and human Lgnl/SMA
RT interval."
RL Genomics 60:137-151(1999).
RN [3]
RP SEQUENCE OF 82-168 FROM N.A.
RC STRAIN=129/SV;
RX MEDLINE=97131520; PubMed=8975718;
RA Scharf J.M., Damron D., Frisella A., Bruno S., Beggs A.H.,
RA Kunkel L.M., Dietrich W.F.;
RT "The mouse region syntenic for human spinal muscular atrophy lies
RT within the Lgnl critical interval and contains multiple copies of Naip
RT exon 5."
RL Genomics 38:405-417(1996).
RC Lgnl
RL FUNCTION: PREVENTS MOTOR-NEURON APOPTOSIS INDUCED BY A VARIETY OF
CC SIGNALS.
CC - SIMILARITY: CONTAINS 3 BIR REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF135492; AAD56764.1; -
CC DR EMBL: AF131205; AAD56760.1; -
CC DR EMBL: U66326; AAC52974.1; -
CC MGD; MGI:1298220; Bircie.

```



```

DR InterPro: IPR001370; BIR.
DR Pfam: PF00653; BIR; 3.
DR SMART: SM00238; BIR; 3.
DR PROSITE: PS01282; BIR_REPEAT_1; 2.
DR PROSITE: PS01423; BIR_REPEAT_2; 3.
DR Apoptosis: Repeat: Multigene family.
KM REPEAT 60 127 BIR 1.
FT REPEAT 159 227 BIR 2.
FT REPEAT 278 345 BIR 3.
FT REPEAT 92 144 S -> R (IN REF. 1).
FT CONFLICT 144 144 S -> R (IN REF. 1).
FT CONFLICT 242 242 S -> G (IN REF. 2).
FT CONFLICT 472 472 T -> A (IN REF. 2).
FT CONFLICT 516 516 A -> D (IN REF. 2).
FT CONFLICT 521 521 A -> T (IN REF. 2).
FT CONFLICT 533 533 V -> A (IN REF. 2).
FT CONFLICT 538 538 S -> I (IN REF. 2).
FT CONFLICT 1092 1092 E -> D (IN REF. 2).
FT CONFLICT 1129 1129 H -> L (IN REF. 2).
FT CONFLICT 1137 1137 R -> Q (IN REF. 2).
FT CONFLICT 1242 1242 V -> I (IN REF. 2).
FT CONFLICT 1276 1276 D -> N (IN REF. 2).
SO SEQUENCE 1403 AA; 159695 MW; B27F645043BCBCE42 CRC64;

```

```

Query Match 12.1%; Score 637.5; DB 1; Length 1403;
Best Local Similarity 23.8%; Pred. No. 5.3e-29;
Matches 248; Conservative 191; Mismatches 407; Indels 195; Gaps 38;

```

```

OY 101 LAQDKLHYHPSFLNF-----YPLGEDIDIIIFNLKSTFTEPVLMRKDOHHR 148
DB 403 LSEQLRDVNTKATFRHMLNPEVCSLSIGDHLSCDVSI--SKHSDPV----- 449
OY 149 VEQLTLNGLALQSPCLIEGSGKSTLQRIAMLGSGKAKLTQKVFYFRLS-- 206
DB 450 GEAULTPEVSNLNSVMCEGETSGKTFKRIAFIVASGCCPLLYRQVLYLSLSI 509
OY 207 RAQGLFETLDQLDIFGTIRKQTFMMLKLRORVLFLLDGYNEFK--PONCEIEAL 264
DB 510 TPDOGLANICQQLAGAGCISEVCLSSIQQLQHOVFLFLDDYGLASLPQ--ALHRL 566
OY 265 IKENHFRNMVIVTTTTCRLRHIRQFALTAEVGDMTEDSAQALIREVLIKEL--AEGIL 322
DB 567 ITKNLSRTCLLIHAVTNKRDIRLYLGTSLIEQEPFNVTAVSRKFFSHDIICVEKLI 626
OY 323 IQIQSRCLRNLMKTPPLFVITCA--IQMGSEFHSHTQTLFHFHYDLLQKNHKKHG 380
DB 627 IYFIDNKDLQGYKTPLEVAAYCTDIQNASAO--DKFQDVTLFQSYMOYLSLKYK---- 680
OY 381 VAASDFIRSLDHCGLALEGVSHKDFEELQV--SSYNEVDVLTGTLCKYTAQRPKR 438
DB 681 ATAERLQATVSSCGDLATGLFSSCFERNSDDLAAGVDEDEKLTLLMSKTTAQRLRPV 740
OY 439 YKFEHKSQOETAGRRSLSLTSHEPEEYVKGNGYLOKMWISIDITSYSSLLRYTCG-S 497
DB 741 YRFLPLPFOELFAAAYRTELISLSDROEDQDGLYLRQIDSPKAIANSFNIPLEYYVSSH 800
OY 498 SVEATRAVMKHLAAVYQCGCLLSIAKRPLMRQESLOS VKTTTQELIKANINISFV-- 555
DB 801 SSKAAPTVVSHL-----IQLVDEKESLEMSNENEYMKLHPOTFLWF 842
OY 556 --ECGIHYOESTSKALSQE-----FEA-----FPOGSLYINSGNIP 592
DB 843 QFVRGLIMLVSPSSSSSFVSEHLRLALIFAYESNVAACSPPILOFLAGKTLALAVLN-- 901
OY 593 DYLDFEFHLPNCASALDPFKIDFYGGAMASWEKAEDTGTGIMEADETYIPSAVSLF 652
DB 902 ---OYFRDHPESSLILSLKVSINGNMSSYVDYSFPT--YFENLQPAIDDEYTSAF 954
OY 653 ---FMMKQEPRLTEVTLDESS-----KINKOD----- 676
DB 955 EHISEWRKRFADDEETIKNYENIRPRALPDISEGYWKLSPKPKCKIPKLEVOVNTDADQ 1014

```

```

OY 677 --ITYLGKIFSSATSRLQIKRCAGVAGSLSVLSTCK--NIYSLWEASPLTFEDERHIT 733
DB 1015 ALLQVLMVEFSAQSQSIIEFRLPMSGSELSICPALDELSKASVYKCSMSRLSLRAQELL 1074
OY 734 SVYTNKTVSI-----HDIQN-----ORLPQG-----LTDISGLNKLTKL 768
DB 1075 TLPALOSLEVSFTNQLPEQLFHNHKLFLGLKELCYRLDQKPDVLSVLPQEPFNHMEKL 1134
OY 769 IMDNKNMNEDEAIKLABGKLNKCKPLRHLTLSDIGEMDYIVASLSSEPC-DIEEIQ 827
DB 1135 SIRT--STESDSKLVKFKLPQNPDLHVFHLK--CDPLSNCESLMAVLAS--CKKREIEF 1188
OY 828 VSCCLSA--NAVKITLQNLNHLNKLISLDL--SENYLEKQENALHELIDPMNVLEQTLALM 885
DB 1189 SRCFEAMFVNIPL-----NFYSKITLNKQDQFPDKETSEKFAQALSLRLEE---LL 1241
OY 886 LPMGCDVQGSLSLKLHLEVPOLVKGKLMRLDTETIRL----GAFGKNPLKNQ 940
DB 1242 VPTGDIHQVAKLIVRQCQLTEPLRVLTFHDIIDDDSVLEIRAMTSGF-----QKLE 1295
OY 941 QNLINAGN--RVSSDGLAFMGVFNELKOLVFPFSTKEFLP-----DPALYKRLSQVLSKL 994
DB 1296 NLDISMNHKITEGYRNFQALDNLPLNQ--ELNLCRNIPGRIOVATTVKALGQCVSRL 1353
OY 995 TFLQEARLVGMQFDDDLVSI 1015
DB 1354 PSLIRLHMISMLLDEEDMKVI 1374

```

RESULT 5

```

BIRB_MOUSE STANDARD: PRT: 1447 AA.

```

```

AC 090UK4: Q9R030: Q09124;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE BACULOVIRAL IAP REPEAT-CONTAINING PROTEIN 1B (NEURONAL APOPTOSIS
DE INHIBITORY PROTEIN 2).
CN BIRCB OR NAIP2 OR NAIP-R56.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99431676; PubMed=10501978;
RA Huang S., Scharf J.M., Growney J.D., Endrizzi M.G., Dietrich W.F.;
RT "The mouse Naip gene cluster on Chromosome 13 encodes several distinct
RT functional transcripts."
RL Mamm. Genome 10:1032-1035(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99315342; PubMed=10384056;
RA Yaraqhi Z., Diez E., Gros P., Mackenzie A.;
RT "cDNA cloning and the 5'genomic organization of Naip2, a candidate
RT gene for murine Legionella resistance."
RL Mamm. Genome 10:761-763(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RX MEDLINE=99417674; PubMed=10486205;
RA Endrizzi M., Huang S., Scharf J.M., Kelter A.R., Wirth B.,
RA Kunkel L.M., Miller W., Dietrich W.F.;
RT "Comparative sequence analysis of the mouse and human Lgn1/SMA
RT interval."
RL Genomics 60:137-151(1999).
RN [4]
RP SEQUENCE OF 82-168 FROM N.A.
RC STRAIN=129/SV;
RX MEDLINE=97131520; PubMed=8975718;
RA Scharf J.M., Damron D., Fritsella A., Bruno S., Beggs A.H.,
RA Kunkel L.M., Dietrich W.F.;
RT "The mouse region syntenic for human spinal muscular atrophy lies

```

RT within the 1gnt critical interval and contains multiple copies of Nalp
 RT exon 5.
 RL Genomics 38:405-417(1996).
 CC - FUNCTION: PREVENTS MOTOR-NEURON APOPTOSIS INDUCED BY A VARIETY OF
 CC SIGNALS.
 CC - SIMILARITY: CONTAINS 3 BIR REPEATS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb.slb.ch).
 CC
 CC EMBL: AF135489; AAD56761.1; -
 CC EMBL: AF135490; AAD56762.1; -
 CC EMBL: AF102871; AAC73002.1; -
 CC EMBL: AF131205; AAD56759.1; -
 CC EMBL: U66329; AAC52977.1; -
 CC MGD: MGI:1298226; Birc1b.
 CC InterPro: IPR001370; BIR.
 CC Pfam: PF00653; BIR; 3.
 CC SMART: SM00236; BIR; 3.
 CC PROSITE: PS01282; BIR_REPEAT_1; 2.
 CC PROSITE: PS0143; BIR_REPEAT_2; 3.
 CC K1 Apoptosis; Repeat; Multigene family.
 CC FT REPEAT 159 127 BIR 1.
 CC FT REPEAT 60 127 BIR 1.
 CC FT REPEAT 159 227 BIR 2.
 CC FT REPEAT 278 345 BIR 3.
 CC FT REPEAT 377 377 BIR 3.
 CC FT CONFLICT 403 403 L -> G (IN REF. 3).
 CC FT CONFLICT 478 478 L -> I (IN REF. 3).
 CC FT CONFLICT 540 540 N -> Y (IN REF. 3).
 CC FT CONFLICT 862 862 K -> N (IN REF. 3).
 CC FT CONFLICT 1079 1080 SD -> FN (IN REF. 3).
 CC FT CONFLICT 1089 1089 R -> C (IN REF. 3).
 CC FT CONFLICT 1115 1115 K -> E (IN REF. 3).
 CC FT CONFLICT 1122 1122 T -> A (IN REF. 3).
 CC FT CONFLICT 1136 1136 D -> E (IN REF. 3).
 CC FT CONFLICT 1157 1157 S -> G (IN REF. 3).
 CC FT CONFLICT 1167 1167 G -> R (IN REF. 3).
 CC FT CONFLICT 1271 1271 F -> C (IN REF. 3).
 CC FT CONFLICT 1447 1447 F -> C (IN REF. 3).
 CC SQ SEQUENCE 1447 AA; 164033 MW; 9EFC6A73BA60A2 CRC64;

Query Match 12.0%; Score 632.5; DB 1; Length 1447;
 Best Local Similarity 24.2%; Pred. No. 1.1e-28;
 Matches 268; Conservative 195; Mismatches 423; Indels 221; Gaps 44;

QY 38 VNICEKGVODARGLIHMILKKGSECNLFKSLKEMNYPLEFDLNGSLFHOTSGD 97
 DB 404 VSYLCNRQD-DHSEAGS-----RGCCASSGTYLPS-----TDL-GGSEQWMLDEA- 444
 QY 98 LDDLAODKLQYHTPSPFLNF-----YPLGEDDIIFNLKSTTEPVLAKRQD 145
 DB 445 -RLSEQLRDTYKATFRHNNLPEVYSLSLGTDLHLSGVII---SKHISQPV----- 493
 QY 146 HHNEVDLTGLGLAOLSPGILGEGSKGKSTLLQRIAMGSKKALKKFFVFLRL 205
 DB 494 ---GGSTLPEVSNLSNVSCVEGASGKTYPLKRIAFWASCCCLNRFQDLVFLSL 550
 QY 206 SRAAGG--LEPTECDLDDIPGTRKOTPMAMLLKRLQVLFLLDGNENK--PONCEI 261
 DB 551 SLLTPGELAKITIAOLGLAGGCISEVCLSSITIQDLOHQLVFLDDYSGLASLPQ---AL 607
 QY 262 EALIKENHRKNNVIYTTTECLRHTRQFAGALAEVGMTEDSQAOLIREVLLKEIAEGL 321
 DB 608 HTLITKNVLSRTCLLAVHTNKVGRIPYLDTSLEIKFEPYTVVSLRKLESHD----- 662
 QY 322 LLDIQR-----SRCLRNKATPLPVYITCALQMGSESESHQTLLFTFDLLIQKN 374
 DB 663 IMVRKFINFGFHEELQGIHKTPLFVAACDTWFKNPSDQPPDVALFKRAYMOYL----- 718

QY 375 KHKHGVASDFIRSLDHCGLALEGVSHKPFDELDV--SSVNEDVLTGLCKYTA 432
 DB 719 SLKHKG-AAKPLQATVSSCCGLALGTGLFSSCFEPNSONLAEAGVDEDELTGLMSKFTA 777
 QY 433 QREPKYKFFHKSFOEYTAGRLSLTSHPEBEVTKGNCYLOKMWISIDITYSLSLR 492
 DB 778 QRUPRYRFGPPLFOELAVRLTELLSDROEDODGLTYLLRQINSPLKAMSIYHTFLK 837
 QY 493 YTGQ-SSVEATRAVMKMLAVVQHGLGLSLAKRPLMRQESLOSQKNTQEOELKAINI 551
 DB 838 YVSHSSKKAAPVYVSHL-----LQLYDEKESLENNSENDYKILAP 879
 QY 552 NS--FVDCGITHLYO-----ESTSKALSQEPAPFQGRSLYI 586
 DB 880 EALLMEICLGLNGWLSPESPLFISENLRLICLNFHESNTVAACSVILQFLRGRTLDL 939
 QY 587 NSGNIPDYLFDFEHLNCAALDFIKLDYGGAMA-----SWEKA----- 627
 DB 940 KVLST-QY--FMDH-PETLLLLKSIKISLNGNMWQRIQFSLIEKSFKEVQPPITDQY 994
 QY 628 -----AEDTGIGHMEAPETIYIPSAVSLFENWK-----QEFRTLEVTLRD 668
 DB 995 AIAFPQINEYQKLNSEKKAHITKKEDEMKHQLPLNISTGY--WKLSKPKYKIPPLEVQVYN 1052
 QY 669 FSKLNKODITYLGKIRSSATSLRLQIRKACAGVAGSLVLTSTCK-NIYSLMVEASPLTIE 727
 DB 1053 TGPADQALLQVLMNEVPSASQIEFRLSDSGPLESIRPALSEKASVTKCSMRLELSRE 1112
 QY 728 DERHITSVNLKLTSLHDLONRPLGGTDSLGNLKNLTYLI----- 769
 DB 1113 DQKLLTLPLTOSLEVS--ETNQLPDLFPHNLKFTLQKELCYRLDSKRPVLSVLPGEFP 1170
 QY 770 ----MDNIMK--NEEDATLAGLKNLKKMCLFHL--THLSDIGEMDYVKSLSSEPC 820
 DB 1171 NLHMEKLSIRSTESDLSLVKILQNSPNLHVFLHLCNPLSC--EPLMTVLASCK-- 1226
 QY 821 DLEIQLVSCCLSA-NAVKILAOENLNLVLTSLDL--SENYLKEGNEALHELIDRMNVL 878
 DB 1227 -LREIFSGRCFEMFVNILP-----NFVPLKILNLMDQGFPPKETSSEKPAQALGSLRNL 1281
 QY 879 EQLTALMLPMGCVQGSLSLKLHLEVPQVYKLGKNNMLTTEIRIL-GAPFGKNPLK 937
 DB 1282 EK--LEFPTGDDIGHQYAKLIVRQCQLPCLRLVLAETLDDSVLEIAGARG----- 1333
 QY 938 NFOOL-NL--AGNRVSSDGMFLPMGVFENLKQLVFDFSTKEFLPD-----DALYKLS 988
 DB 1334 GFOKLEMLDITLHKKITEEGYRNFPQVLDNLPNKLNDIS--RHIEPCIOIQAITYKALG 1391
 QY 989 QVLSKITFLQEARLVGNQFDDDLVYI 1015
 DB 1392 QCVSRPLSLTRGLMLSVLDDDEDIKVI 1418

RESULT 6
 BIR_MOUSE
 ID BIR_MOUSE STANDARD; PRT: 1403 AA.
 AC 090KMS; 09R017; 09JIB5;
 DT 20-AUG-2001 (rel. 40, created)
 DT 20-AUG-2001 (rel. 40, last sequence update)
 DT 20-AUG-2001 (rel. 40, last annotation update)
 DE BACULOVIRAL IAP REPEAT-CONTAINING PROTEIN 1A (NEURONAL APOPTOSIS
 DE INHIBITORY PROTEIN 1).
 GN BIR3A OR NAPI1 OR NAPI.
 OS Mus musculus (mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Varaghi Z., Korneljuk R.G., Mackenzie A.E.;
 RT "Cloning and characterization of the multiple copies of the murine
 homolog of Nalp (neuronal apoptosis inhibitory protein).";

```

RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RA SEQUENCE FROM N.A.
RX MEDLINE=99431676; PubMed=10501978;
RA Huang S., Scharf J.M., Growney J.D., Endrizzi M.G., Dietrich W.F.;
RT "The mouse Nalp gene cluster on Chromosome 13 encodes several distinct
RT functional transcripts."
RL Mamm. Genome 10:1032-1035(1999).
RN [3]
RA SEQUENCE FROM N.A.
RX MEDLINE=20411747; PubMed=10958627;
RA Endrizzi M.G., Hadjioio V., Growney J.D., Miller W., Dietrich W.F.;
RT "Genomic sequence analysis of the mouse Nalp gene array."
RL Genome Res. 10:1095-1102(2000).
CC -1- FUNCTION: PREVENTS MOTOR-NEURON APOPTOSIS INDUCED BY A VARIETY OF
CC SIGNALS.
CC -1- SIMILARITY: CONTAINS 3 BIR REPEATS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC
CC EMBL; AF007769; AAB69223.1; -
CC EMBL; AF135491; AAD56753.1; -
CC EMBL; AF242432; AAF82752.1; -
CC MGD; MGI:129823; Birc4.
CC InterPro: IPR001370; BIR.
CC
CC P1am; PF00653; BIR; 3.
CC SMART; SM00238; BIR; 3.
CC PROSITE; PS01282; BIR_REPEAT_1; 1.
CC PROSITE; PS01043; BIR_REPEAT_2; 3.
CC
CC DR Apoptosis; Repeat; Multigene family.
CC KW REPEAT
CC FT REPEAT 60 127 BIR 1.
CC FT REPEAT 159 227 BIR 2.
CC FT REPEAT 278 345 BIR 3.
CC FT REPEAT 343 343 I -> V (IN REF. 2).
CC FT CONFLICT 359 359 L -> O (IN REF. 2).
CC FT CONFLICT 624 624 E -> K (IN REF. 2).
CC FT CONFLICT 1092 1092 D -> E (IN REF. 3).
CC FT CONFLICT 1116 1116 D -> N (IN REF. 3).
CC FT CONFLICT 1123 1123 G -> R (IN REF. 1).
CC FT CONFLICT 1129 1129 L -> H (IN REF. 3).
CC FT CONFLICT 1140 1140 T -> M (IN REF. 2).
CC FT CONFLICT 1269 1269 A -> V (IN REF. 3).
CC
CC FT 1403 AA; 158692 MW; B316302595955EE67 CRC64;
CC SEQUENCE

```

[illegible]

Db	595	TSEIKEERPLSNVTYLLKFKFNSNIKRILE--FWVYEGQMEDLDGIIHKTRPLFAANCTDWE	655
Qy	350	GESEPHSHTOTTFTEHFYDILLQKNKKHKGVASDFIRSLDHCGDLAEGVESHKEDFE	409
Db	654	ENSPDQEPFODMALKSYMOYL-----SLKHG--AAKPLQATVSSCGQALATLGSSCFEEN	708
Qy	410	LQDY--SSVANDVLLTGGGLCKYTAQRFKPKYFHKFSQOETTAGRRSSLTSHREPEV	467
Db	709	SDDLAEGVDEDELTTLCKMSKFTPAQRLREYVFEPLRPFQEFLEAMRLTELLSSDRQDQ	768
Qy	468	TKGNQYLQKMWISLDITSTYSSSLRYTCG--SSVEATRAVMKILAAVYHGCL-----	519
Db	769	DLGLYLRQINSPLKATLTYYNNLKYFHSPPSKAPRYVSHLLHVDETELLETTYNE	828
Qy	520	-----GLS--IAK--RPLMRQESLOSVKNTTEQELKAINI-----NSFVECGIHLVY--	563
Db	829	DYVNHPRGTSTIKMGLKELMLSPREYSSFVSEPHLLRLIALNPAYESNVAECSPRILOFL	888
Qy	564	-----ESTSKSALSQEF-----EAFQGSLL--YNSGNIP---DY--LDFEPEHL-----P	603
Db	889	RGRLAKLVYLNLOQFRDHPSSLVLSVLEYSINGNKVPKVVYDYSWEKSFETLPPTIDQ	948
Qy	604	NCASALDFIKLDFYGMAAMEKAADDTGIIHMEAPETYSRAVYSF-----	653
Db	949	DYSAAFE-----QMKHEK-----NISEEELI-----KIKINIFLOPPKISSG	988
Qy	654	NWKO-----EFTLEVTLPDFSKLNKODITYLOKIFPSATSRLQIKKQAGVAGSLVL	708
Db	989	YKLSRPKPKIPRLVEYGVTMGPRADQALLQVLMVEFVSASQSTIEFLRDSOSGFLEIRPAL	1048
Qy	709	STCK-NIYSLMVEASPLTIDENHITTSVYNKLTLSIHDLOQRPLCGLTSL-----	759
Db	1049	ELSKASVTKCSMSRLELSRAQDELLTLPALOSLEVS--ETNQLRDPOLFHNHMKFLGLKE	1106
Qy	760	-----GNLKNLTKRLIMDNIMKM--NEEDAIKLAGIKNKKMCFHLTH	800
Db	1107	LCVFLDGKPDVLSVLPPEFPNL--LHMEKISTSTESDLSKVLKLIQNSPULHVFHLK-	1163
Qy	801	LSDIGEMDYIVKSLSESPC-DLEIQLVSSCLSANVA-KPIAOWJHNLVKLSIDL-SE	857
Db	1164	-CDPLSNCDSLMVLAS--CKRLREIEFGSRCEAMPFNILP-----NFIKLKILNLSQ	1216
Qy	858	NYLEKQDNEALHEIDRANVLBDQLMLPMGCDVOGSLSLKLYLEVPOLVYIGLKNW	917
Db	1217	QEPKKEISEKFAQALGSLRMLEE--LLVPTGGIHQVAKLIVRCLOLPCILRYLAHYI	1273
Qy	918	RLTTEIRIL-----GAFEFKNPLKNNQOULAGN-RVSSDGLMFMGVENFLKQVDFD	971
Db	1274	LNDNSVEIIRAVATSGCF-----QKLEKLDISMNHKITTEEBYRNFPQALDILPMLQ--N	1325
Qy	972	FSTKEFLPD-----PALVRKLSGVLSKTLFLOEARLVGMQFDDDDLSVI	1015
Db	1326	LNICRHIREPCIOYQATVTKALGCVSRLPSLTPRLHLSMLDDEDMKVI	1374
RESULT 7			
C2PA.MOUSE ID C2PA.MOUSE STANDARD; PRT: 1078 AA.			
AC	P79621: Q31115:		
DT	15-DEC-1998 (Rel. 37, Created)		
DT	15-DEC-1998 (Rel. 37, Last sequence update)		
DT	20-AUG-2001 (Rel. 40, Last annotation update)		
DE	MHC CLASS II TRANSCRIPTIVANOR CITTA.		
GN	MHC2PA OR CITTA OR C2PA.		
OS	Mus musculus (Mouse)		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.		
OX	NCBI_TaxID:10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-NOD: TISSUE-Spleen;		
KX	MEDLINE=97152412; Pubmed=8995190;		
RA	Stins T.N., Elliott J.F., Ramassar V., Denney D.W. Jr., Halloran P.F.;		

RT "Mouse class II transactivator: cDNA sequence and amino acid
RT comparison with the human class II transactivator."
RL Immunogenetics 45:220-222(1997).
RN [2]
RP SEQUENCE OF 878-1020 FROM N.A.
RC STRAIN-BALB/C; TISSUE-Thymus;
RX MEDLINE-96188886; PubMed-8620527;
RA Panelli M.C., Wang E., Shen S., Schluter S.F., Bernstein R.M.,
RA Herish E.M., Stoeck A., Gangavalli R., Barber J., Jolly D.,
RA Akporiaye E.T.;
RT "Interferon gamma (IFNgamma) gene transfer of an EMT6 tumor that is
RT poorly responsive to IFNgamma stimulation: Increase in tumor
RT immunogenicity is accompanied by induction of a mouse class II
RT transactivator and class II MHC."
RL Cancer Immunol. Immunother. 42:99-107(1996).
CC -!- FUNCTION: ESSENTIAL FOR TRANSCRIPTIONAL ACTIVITY OF THE HLA CLASS
CC II PROMOTER: ACTIVATION IS VIA THE PROXIMAL PROMOTER. NO DNA
CC BINDING OF IN VITRO TRANSLATED CITRA WAS DETECTED. MAY ACT IN A
CC COACTIVATOR-LIKE FASHION THROUGH PROTEIN-PROTEIN INTERACTIONS BY
CC CONTACTING FACTORS BINDING TO THE PROXIMAL MHC CLASS II PROMOTER.
CC TO ELEMENTS OF THE TRANSCRIPTION MACHINERY, OR BOTH. ALTERNATIVELY
CC IT MAY ACTIVATE HLA CLASS II TRANSCRIPTION BY MODIFYING PROTEINS
CC THAT BIND TO THE MHC CLASS II PROMOTER (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; 060653; AAB4859.1; -
DR EMBL; 046562; AAB05004.1; -
DR MGI; 108445; C2ta.
DR InterPro: IPR003592; LRR_out.
DR SMART; SM00370; LRR; 4.
KW Transcription regulation; Activator; ATP-binding.
FT DOMAIN 52 138 ASP/GLU-RICH (ACIDIC).
FT NP_BIND 368 375 ATP (POTENTIAL).
FT CONFLICT 917 917 G -> R (IN REF. 2).
SQ SEQUENCE 1078 AA; 118789 MW; 2BFEA61BBF561378 CNC64;

Query Match 3.4%; Score 180.5; DB 1; Length 1078;
Best local similarity 19.7%; Pred. No. 0.0062;
Matches 176; Conservative 126; Mismatches 304; Indels 289; Gaps 42;

QY 151 QLTNLGLQLAQ--SPC-----IIEGSGKSTLLQRIAMLMSGKCAALTEKE 199
DB 339 QLAHGGIAEVLQVSDCRREPGETQYVAVLGAKGQKSHMARVSHWA---CGQLQYDF 395
200 VFELR--LSRAQG--GLFETLC-----DQLDIGTIRKQYFMALRLKROVL 244
DB 396 VFYVPCGCLDRPDYTHLRLLCPRLPSLOPLAMDEVID-----YIAQPRVL 443
QY 245 FLIDGYNFRPQNCPELIALIKENHFRKNVIVTTTECLRHIRQFGALTAEGDMDT- 303
DB 444 LIIDAFEELEAD-----GLLHGPGSLSPEP 470
QY 304 -SQAQILRELIKELAEGL--LLLOQKSRCLNIMKTPLEFVITCLOMGSEFHSHTQ 359
DB 471 CSIRGLAGLFFKRLRGCTLLTPARGRSLAOSKAD-----AIFEVSFSTQAK 523
QY 360 TTFEHTFYDILIOKHKHKGVAASDFIRSLDHCGLALEG-----VFSEKPELQDVS 414
DB 524 TYMRHFFENSGTGNQKAGL-----LEGQPLGCSYSH----- 557
QY 415 SVNEDVLLTTGLCKYTAQFKPKYKFFHKSFOEYTAGRLSSLLTS-----HEPEEYFK 469
DB 558 -----SPVVCRAVQC-----LSKALLEQGEADLPCTLTGLYVSLGPAQONS 600
QY 470 GNGYLOKMWIS-DISTYSSLLRYTCGSSVEA-TRAVMKHLAAVYOHGCLLGIATAKRP 527

DB 601 PCGALVELAKLAWELGRRHSTLOETRFSSVEYKTAWVQGL----- 642
QY 528 LMRQESLOSVKNTQEDELKAININSP-VEC---GIHLVQESNSK-SALSQEFAPFOGK 562
DB 643 -----MOOTLETTAOLA-----FSSFLQCLYGAWLAQCNKIKELFOYALTPRKK 692
QY 583 SLXIN-SGNIPDLDFE-FEHLNPKASAL-----DEFLKDFYGAMAS 623
DB 693 RPYDNMEGVPRLAGVFPRAHCGALVEPAVAADVKKQVLRYLKRLGLGLRAG 752
QY 624 -----WEKAEDTGIMHE-----EAPETVPSRAVSLEFNMKOEFR 660
DB 753 RLLELLHCAHETQPGIWEHVAHQLPg-HLSFLGTRLTLPDYYVLGRA----- 799
QY 661 TLEVTLRDESKLWKO---DIYVIGKI--FESATSLRQIKRCGVAGSL-----SLVLSF 710
DB 800 -LETASODFSLDRQGVESFGLNVLGSLCVTSFRASLSDTMALWESLQOQGEAQLLA 858
QY 711 CKNIYSL--WEASPLTIEDERHITSVNLKITS-----IHDLOQRPLPG----- 754
DB 859 AEKFTIEPKAKSPKQVEDLDRLVQORLRNPSSEDAKQDPAIRDLKLEFALGPLGP 918
QY 755 -----LTDLSGNLKNLFLKIMDN---KNEEDAIKLABGLKMKCLEFHL--HLSDI 804
DB 919 QAFPTLAKLIPAFSSLOHLDLSLENKIDGKVSRLSATFPOLKALETLNLSQNNITDV 978
QY 805 G-----EGMDYIVKSSSEPCDEEIQVSCSANAANKITLONLNLVLTSLDSENY 859
DB 979 GACKLAEALALAKSL-----RSLYNNCICDKCAKSLAQLVDPMSLSLRMDVQFNK 1031
QY 860 LEKQNEALHELIDRNVLEQLT--ALMLFPGCDVQGSLSLLKHLEVPQVLK 912
DB 1032 FTAAGAOQLASSQKCPQVETLAMWPTPIPFg-----VQEHQQLDARISL 1077

RESULT 8
ID AACT_DICDI STANDARD; PRT; 862 AA.
AC P05095;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ALPHA-ACTININ, NON-MUSCULAR (F-ACTIN CROSS LINKING PROTEIN).
GN ABPA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AX2;
RX MEDLINE-87304850; PubMed-3622778;
RA Noegel A., Witke W., Schleicher M.;
RT "Calcium-sensitive non-muscle alpha-actinin contains EF-hand
RT structures and highly conserved regions."
RL FEBS Lett. 221:391-396(1987).
RN [2]
RP SEQUENCE OF 92-505 FROM N.A.
RC STRAIN-AX2;
RX MEDLINE-86304574; PubMed-3745276;
RA Witke W., Schleicher M., Lottspeich F., Noegel A.;
RT "Studies on the transcription, translation, and structure of alpha-
RT actinin in Dictyostelium discoideum."
RL J. Cell Biol. 103:969-975(1986).
CC -!- FUNCTION: F-ACTIN CROSS-LINKING PROTEIN WHICH IS THOUGHT TO ANCHOR
CC ACTIN TO A VARIETY OF INTRACELLULAR STRUCTURES. THIS IS A BUNDLING
CC PROTEIN.
CC -!- SUBUNIT: HOMODIMER, ANTIPARALLEL.
CC -!- SIMILARITY: CONTAINS 1 ACTIN-BINDING DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY DOMAINS (CH).
CC -!- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
CC -!- SIMILARITY: CONTAINS 4 SPECTRIN REPEATS.
CC -----

RT "Characterization of a novel giant scaffolding protein, CG-MAP, that
RT anchors multiple signaling enzymes to centrosome and the golgi
RT apparatus";
RT J. Biol. Chem. 274:17267-17274(1999).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RA Kemmer W.A., Deiss S., Schwarz U.;
RT "Cloning of Hyperion.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 323-391 FROM N.A. (ISOFORM 2).
RA MEDLINE-99115654; PubMed-9915845;
RX Schmidt P.H., Dransfield D.T., Claudio J.O., Hawley R.G.,
RA Trotter K.W., Milgram S.L., Goldenring J.R.;
RT "AKAP350, a multiply spliced protein kinase A-anchoring protein
RT associated with centrosomes";
RL J. Biol. Chem. 274:3055-3066(1999).
RN [6]
RP SEQUENCE OF 1802-3876 FROM N.A. (ISOFORM 5).
RA TISSUE-Lymphoblast;
RX Hinds K., Sutterer C., Becker M., Hawkins M.;
RT Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE OF 2157-3911 FROM N.A. (ISOFORM 6).
RA TISSUE-Lung;
RX Milgram S.L., Goldenring J.R., Schmidt P.H.;
RT "AKAP350: A multiply spliced family of proteins with centrosomal
RT association.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [8]
RP SEQUENCE OF 2212-3911 FROM N.A. (ISOFORM 2/3).
RA TISSUE-Brain;
RX MEDLINE-99087487; PubMed-9872452;
RX Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XI.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 5:277-286(1998).
RN [9]
RP SEQUENCE OF 17-1800 FROM N.A.
RA Wu X., Graves T., Bradshaw H.;
RT Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [10]
RP FUNCTION: BINDS TO TYPE II REGULATORY SUBUNITS OF PROTEIN KINASE
CC A. SCAFFOLDING PROTEIN THAT ASSEMBLES SEVERAL PROTEIN KINASES AND
CC PHOSPHATASES ON CENTROSOME AND GOLGI APPARATUS WHERE PHYSIOLOGICAL
CC EVENTS CAN BE REGULATED BY PHOSPHORYLATION STATE OF PROTEIN
CC SUBSTRATES. ISOFORM 4/YOTIAO IS ASSOCIATED WITH THE N-METHYL-D-
CC ASPARTATE RECEPTOR AND IS SPECIFICALLY FOUND IN THE NEUROMUSCULAR
CC JUNCTION (NMJ) AS WELL AS IN NEURONAL SYNAPSES EXPLAINING THAT ITS
CC ROLE MAY BE TO ORGANIZE POSTSYNAPTIC SPECIALIZATIONS.
CC SUBUNIT: INTERACTS WITH THE REGULATORY REGION OF PROTEIN KINASE N
CC (PKN), PROTEIN PHOSPHATASE 2A (PP2A), PROTEIN PHOSPHATASE 1 (Pp1)
CC AND THE IMMATURE NON-PHOSPHORYLATED FORM OF PKC EPSILON.
CC SUBCELLULAR LOCATION: CENTROSOMAL IN MANY CELL TYPES AND
CC CYTOSOL IN PARIENTAL CELLS.
CC -1- ALTERNATIVE PRODUCTS: 6 ISOFORMS; 1 (SHOWN HERE), 2, 3/CG-
CC NAP, 4/YOTIAO, 5 AND 6/AKAP350, ARE PRODUCED BY ALTERNATIVE
CC SPLICING.
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED. ISOFORM 4/YOTIAO IS HIGHLY
CC EXPRESSED IN SKELETAL MUSCLE AND IN PANCREAS.
CC -1- DOMAIN: RII BINDING SITE, PREDICTED TO FORM AN AMPHIPATHIC HELIX,
CC COULD PARTICIPATE IN PROTEIN-PROTEIN INTERACTIONS WITH A
CC COMPLEMENTARY SURFACE ON THE R-SUBUNIT DIMER.
CC -1- CAUTION: REF. 6 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO TWO
CC FRAMESHIFTS IN POSITIONS 3782 AND 3811.
CC -1- CAUTION: REF. 9 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO FOUR
CC FRAMESHIFTS IN POSITIONS 29, 1653, 1699 AND 1735.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AJ131693; CAB40713.1; -
DR EMBL: AB019691; BAA78718.1; -
DR EMBL: AF010770; CAA09361.1; -
DR EMBL: AF026245; AAB86384.1; -
DR EMBL: AF083037; AAD22767.1; -
DR EMBL: AF004013; AAB96867.1; -
DR EMBL: AF091711; AAD39719.1; -
DR EMBL: AB018346; BAA34523.1; -
DR EMBL: AC000066; AAC60380.1; ALT_FRAME.
DR MIM: 604001; -
KW Coiled coil; Alternative splicing; Polymorphism.
FT DOMAIN 2554 2567 PKA-RII SUBUNIT BINDING DOMAIN.
FT 164 914
FT DOMAIN 944 1022
FT DOMAIN 1100 1185
FT DOMAIN 1253 1280
FT DOMAIN 1336 1392
FT DOMAIN 1434 1459
FT DOMAIN 1585 1659
FT DOMAIN 1857 2435
FT DOMAIN 2544 2561
FT DOMAIN 2603 2776
FT DOMAIN 3065 3092
FT DOMAIN 3124 3470
FT DOMAIN 3587 3689
FT DOMAIN 3726 3730
FT DOMAIN 203 292
FT DOMAIN 321 1010
FT DOMAIN 1846 2772
FT VARSPLIC 17 28
FT VARSPLIC 1637 1642
FT VARSPLIC 1643 3911
FT VARSPLIC 2175 2183
FT VARSPLIC 2895 2907
FT VARSPLIC 2895 2948
FT VARSPLIC 3901 3911
FT VARIANT 1347 1347
FT 76 76
FT CONFLICT 475 475
FT CONFLICT 554 554
FT CONFLICT 638 638
FT CONFLICT 663 663
FT CONFLICT 913 913
FT CONFLICT 956 956
FT CONFLICT 980 982
FT CONFLICT 997 997
FT CONFLICT 1001 1001
FT CONFLICT 1020 1020
FT CONFLICT 1028 1028
FT CONFLICT 1626 1626
FT CONFLICT 1703 1703
FT CONFLICT 1707 1707
FT CONFLICT 1802 1803
FT CONFLICT 1843 1843
FT CONFLICT 1956 1956
FT CONFLICT 2027 2027
FT CONFLICT 2157 2158
FT CONFLICT 2169 2169
FT CONFLICT 2514 2514
FT CONFLICT 2851 2851
FT CONFLICT 2957 2957
FT CONFLICT 2983 2983
FT CONFLICT 3087 3087
FT CONFLICT 3218 3218
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
K -> KO.
/FTID=VAR.010926.
E -> Q (IN REF. 3).
M -> I (IN REF. 3).
E -> G (IN REF. 3).
R -> S (IN REF. 3).
N -> S (IN REF. 3).
H -> N (IN REF. 3).
K -> N (IN REF. 3).
OKH -> PKP (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
N -> D (IN REF. 3).
V -> E (IN REF. 3).
R -> P (IN REF. 1 AND 2).
N -> T (IN REF. 3).
V -> G (IN REF. 3).
MISSING (IN REF. 5).
A -> P (IN REF. 3).
I -> V (IN REF. 3).
V -> D (IN REF. 5).
EI -> HE (IN REF. 7).
E -> V (IN REF. 3).
L -> R (IN REF. 3).
I -> N (IN REF. 8).
E -> D (IN REF. 3).
P -> S (IN REF. 3).
Q -> H (IN REF. 3).
Q -> H (IN REF. 3).
STTQFHAGMR -> ALSITTSQHSAPTAPELITLS
SLG (IN ISOFORM 6).
K -> KO.
/FTID=VAR.010926.
E -> Q (IN REF. 3).
M -> I (IN REF. 3).
E -> G (IN REF. 3).
R -> S (IN REF. 3).
N -> S (IN REF. 3).
H -> N (IN REF. 3).
K -> N (IN REF. 3).
OKH -> PKP (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
N -> D (IN REF. 3).
V -> E (IN REF. 3).
R -> P (IN REF. 1 AND 2).
N -> T (IN REF. 3).
V -> G (IN REF. 3).
MISSING (IN REF. 5).
A -> P (IN REF. 3).
I -> V (IN REF. 3).
V -> D (IN REF. 5).
EI -> HE (IN REF. 7).
E -> V (IN REF. 3).
L -> R (IN REF. 3).
I -> N (IN REF. 8).
E -> D (IN REF. 3).
P -> S (IN REF. 3).
Q -> H (IN REF. 3).
Q -> H (IN REF. 3).
STTQFHAGMR -> ALSITTSQHSAPTAPELITLS
SLG (IN ISOFORM 6).
K -> KO.
/FTID=VAR.010926.
E -> Q (IN REF. 3).
M -> I (IN REF. 3).
E -> G (IN REF. 3).
R -> S (IN REF. 3).
N -> S (IN REF. 3).
H -> N (IN REF. 3).
K -> N (IN REF. 3).
OKH -> PKP (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
N -> D (IN REF. 3).
V -> E (IN REF. 3).
R -> P (IN REF. 1 AND 2).
N -> T (IN REF. 3).
V -> G (IN REF. 3).
MISSING (IN REF. 5).
A -> P (IN REF. 3).
I -> V (IN REF. 3).
V -> D (IN REF. 5).
EI -> HE (IN REF. 7).
E -> V (IN REF. 3).
L -> R (IN REF. 3).
I -> N (IN REF. 8).
E -> D (IN REF. 3).
P -> S (IN REF. 3).
Q -> H (IN REF. 3).
Q -> H (IN REF. 3).
STTQFHAGMR -> ALSITTSQHSAPTAPELITLS
SLG (IN ISOFORM 6).
K -> KO.
/FTID=VAR.010926.
E -> Q (IN REF. 3).
M -> I (IN REF. 3).
E -> G (IN REF. 3).
R -> S (IN REF. 3).
N -> S (IN REF. 3).
H -> N (IN REF. 3).
K -> N (IN REF. 3).
OKH -> PKP (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
N -> D (IN REF. 3).
V -> E (IN REF. 3).
R -> P (IN REF. 1 AND 2).
N -> T (IN REF. 3).
V -> G (IN REF. 3).
MISSING (IN REF. 5).
A -> P (IN REF. 3).
I -> V (IN REF. 3).
V -> D (IN REF. 5).
EI -> HE (IN REF. 7).
E -> V (IN REF. 3).
L -> R (IN REF. 3).
I -> N (IN REF. 8).
E -> D (IN REF. 3).
P -> S (IN REF. 3).
Q -> H (IN REF. 3).
Q -> H (IN REF. 3).
STTQFHAGMR -> ALSITTSQHSAPTAPELITLS
SLG (IN ISOFORM 6).
K -> KO.
/FTID=VAR.010926.
E -> Q (IN REF. 3).
M -> I (IN REF. 3).
E -> G (IN REF. 3).
R -> S (IN REF. 3).
N -> S (IN REF. 3).
H -> N (IN REF. 3).
K -> N (IN REF. 3).
OKH -> PKP (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
N -> D (IN REF. 3).
V -> E (IN REF. 3).
R -> P (IN REF. 1 AND 2).
N -> T (IN REF. 3).
V -> G (IN REF. 3).
MISSING (IN REF. 5).
A -> P (IN REF. 3).
I -> V (IN REF. 3).
V -> D (IN REF. 5).
EI -> HE (IN REF. 7).
E -> V (IN REF. 3).
L -> R (IN REF. 3).
I -> N (IN REF. 8).
E -> D (IN REF. 3).
P -> S (IN REF. 3).
Q -> H (IN REF. 3).
Q -> H (IN REF. 3).
STTQFHAGMR -> ALSITTSQHSAPTAPELITLS
SLG (IN ISOFORM 6).
K -> KO.
/FTID=VAR.010926.
E -> Q (IN REF. 3).
M -> I (IN REF. 3).
E -> G (IN REF. 3).
R -> S (IN REF. 3).
N -> S (IN REF. 3).
H -> N (IN REF. 3).
K -> N (IN REF. 3).
OKH -> PKP (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
N -> D (IN REF. 3).
V -> E (IN REF. 3).
R -> P (IN REF. 1 AND 2).
N -> T (IN REF. 3).
V -> G (IN REF. 3).
MISSING (IN REF. 5).
A -> P (IN REF. 3).
I -> V (IN REF. 3).
V -> D (IN REF. 5).
EI -> HE (IN REF. 7).
E -> V (IN REF. 3).
L -> R (IN REF. 3).
I -> N (IN REF. 8).
E -> D (IN REF. 3).
P -> S (IN REF. 3).
Q -> H (IN REF. 3).
Q -> H (IN REF. 3).
STTQFHAGMR -> ALSITTSQHSAPTAPELITLS
SLG (IN ISOFORM 6).
K -> KO.
/FTID=VAR.010926.
E -> Q (IN REF. 3).
M -> I (IN REF. 3).
E -> G (IN REF. 3).
R -> S (IN REF. 3).
N -> S (IN REF. 3).
H -> N (IN REF. 3).
K -> N (IN REF. 3).
OKH -> PKP (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
N -> D (IN REF. 3).
V -> E (IN REF. 3).
R -> P (IN REF. 1 AND 2).
N -> T (IN REF. 3).
V -> G (IN REF. 3).
MISSING (IN REF. 5).
A -> P (IN REF. 3).
I -> V (IN REF. 3).
V -> D (IN REF. 5).
EI -> HE (IN REF. 7).
E -> V (IN REF. 3).
L -> R (IN REF. 3).
I -> N (IN REF. 8).
E -> D (IN REF. 3).
P -> S (IN REF. 3).
Q -> H (IN REF. 3).
Q -> H (IN REF. 3).
STTQFHAGMR -> ALSITTSQHSAPTAPELITLS
SLG (IN ISOFORM 6).
K -> KO.
/FTID=VAR.010926.
E -> Q (IN REF. 3).
M -> I (IN REF. 3).
E -> G (IN REF. 3).
R -> S (IN REF. 3).
N -> S (IN REF. 3).
H -> N (IN REF. 3).
K -> N (IN REF. 3).
OKH -> PKP (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
N -> D (IN REF. 3).
V -> E (IN REF. 3).
R -> P (IN REF. 1 AND 2).
N -> T (IN REF. 3).
V -> G (IN REF. 3).
MISSING (IN REF. 5).
A -> P (IN REF. 3).
I -> V (IN REF. 3).
V -> D (IN REF. 5).
EI -> HE (IN REF. 7).
E -> V (IN REF. 3).
L -> R (IN REF. 3).
I -> N (IN REF. 8).
E -> D (IN REF. 3).
P -> S (IN REF. 3).
Q -> H (IN REF. 3).
Q -> H (IN REF. 3).
STTQFHAGMR -> ALSITTSQHSAPTAPELITLS
SLG (IN ISOFORM 6).
K -> KO.
/FTID=VAR.010926.
E -> Q (IN REF. 3).
M -> I (IN REF. 3).
E -> G (IN REF. 3).
R -> S (IN REF. 3).
N -> S (IN REF. 3).
H -> N (IN REF. 3).
K -> N (IN REF. 3).
OKH -> PKP (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
N -> D (IN REF. 3).
V -> E (IN REF. 3).
R -> P (IN REF. 1 AND 2).
N -> T (IN REF. 3).
V -> G (IN REF. 3).
MISSING (IN REF. 5).
A -> P (IN REF. 3).
I -> V (IN REF. 3).
V -> D (IN REF. 5).
EI -> HE (IN REF. 7).
E -> V (IN REF. 3).
L -> R (IN REF. 3).
I -> N (IN REF. 8).
E -> D (IN REF. 3).
P -> S (IN REF. 3).
Q -> H (IN REF. 3).
Q -> H (IN REF. 3).
STTQFHAGMR -> ALSITTSQHSAPTAPELITLS
SLG (IN ISOFORM 6).
K -> KO.
/FTID=VAR.010926.
E -> Q (IN REF. 3).
M -> I (IN REF. 3).
E -> G (IN REF. 3).
R -> S (IN REF. 3).
N -> S (IN REF. 3).
H -> N (IN REF. 3).
K -> N (IN REF. 3).
OKH -> PKP (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
N -> D (IN REF. 3).
V -> E (IN REF. 3).
R -> P (IN REF. 1 AND 2).
N -> T (IN REF. 3).
V -> G (IN REF. 3).
MISSING (IN REF. 5).
A -> P (IN REF. 3).
I -> V (IN REF. 3).
V -> D (IN REF. 5).
EI -> HE (IN REF. 7).
E -> V (IN REF. 3).
L -> R (IN REF. 3).
I -> N (IN REF. 8).
E -> D (IN REF. 3).
P -> S (IN REF. 3).
Q -> H (IN REF. 3).
Q -> H (IN REF. 3).
STTQFHAGMR -> ALSITTSQHSAPTAPELITLS
SLG (IN ISOFORM 6).
K -> KO.
/FTID=VAR.010926.
E -> Q (IN REF. 3).
M -> I (IN REF. 3).
E -> G (IN REF. 3).
R -> S (IN REF. 3).
N -> S (IN REF. 3).
H -> N (IN REF. 3).
K -> N (IN REF. 3).
OKH -> PKP (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
N -> D (IN REF. 3).
V -> E (IN REF. 3).
R -> P (IN REF. 1 AND 2).
N -> T (IN REF. 3).
V -> G (IN REF. 3).
MISSING (IN REF. 5).
A -> P (IN REF. 3).
I -> V (IN REF. 3).
V -> D (IN REF. 5).
EI -> HE (IN REF. 7).
E -> V (IN REF. 3).
L -> R (IN REF. 3).
I -> N (IN REF. 8).
E -> D (IN REF. 3).
P -> S (IN REF. 3).
Q -> H (IN REF. 3).
Q -> H (IN REF. 3).
STTQFHAGMR -> ALSITTSQHSAPTAPELITLS
SLG (IN ISOFORM 6).
K -> KO.
/FTID=VAR.010926.
E -> Q (IN REF. 3).
M -> I (IN REF. 3).
E -> G (IN REF. 3).
R -> S (IN REF. 3).
N -> S (IN REF. 3).
H -> N (IN REF. 3).
K -> N (IN REF. 3).
OKH -> PKP (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
N -> D (IN REF. 3).
V -> E (IN REF. 3).
R -> P (IN REF. 1 AND 2).
N -> T (IN REF. 3).
V -> G (IN REF. 3).
MISSING (IN REF. 5).
A -> P (IN REF. 3).
I -> V (IN REF. 3).
V -> D (IN REF. 5).
EI -> HE (IN REF. 7).
E -> V (IN REF. 3).
L -> R (IN REF. 3).
I -> N (IN REF. 8).
E -> D (IN REF. 3).
P -> S (IN REF. 3).
Q -> H (IN REF. 3).
Q -> H (IN REF. 3).
STTQFHAGMR -> ALSITTSQHSAPTAPELITLS
SLG (IN ISOFORM 6).
K -> KO.
/FTID=VAR.010926.
E -> Q (IN REF. 3).
M -> I (IN REF. 3).
E -> G (IN REF. 3).
R -> S (IN REF. 3).
N -> S (IN REF. 3).
H -> N (IN REF. 3).
K -> N (IN REF. 3).
OKH -> PKP (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
N -> D (IN REF. 3).
V -> E (IN REF. 3).
R -> P (IN REF. 1 AND 2).
N -> T (IN REF. 3).
V -> G (IN REF. 3).
MISSING (IN REF. 5).
A -> P (IN REF. 3).
I -> V (IN REF. 3).
V -> D (IN REF. 5).
EI -> HE (IN REF. 7).
E -> V (IN REF. 3).
L -> R (IN REF. 3).
I -> N (IN REF. 8).
E -> D (IN REF. 3).
P -> S (IN REF. 3).
Q -> H (IN REF. 3).
Q -> H (IN REF. 3).
STTQFHAGMR -> ALSITTSQHSAPTAPELITLS
SLG (IN ISOFORM 6).
K -> KO.
/FTID=VAR.010926.
E -> Q (IN REF. 3).
M -> I (IN REF. 3).
E -> G (IN REF. 3).
R -> S (IN REF. 3).
N -> S (IN REF. 3).
H -> N (IN REF. 3).
K -> N (IN REF. 3).
OKH -> PKP (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
N -> D (IN REF. 3).
V -> E (IN REF. 3).
R -> P (IN REF. 1 AND 2).
N -> T (IN REF. 3).
V -> G (IN REF. 3).
MISSING (IN REF. 5).
A -> P (IN REF. 3).
I -> V (IN REF. 3).
V -> D (IN REF. 5).
EI -> HE (IN REF. 7).
E -> V (IN REF. 3).
L -> R (IN REF. 3).
I -> N (IN REF. 8).
E -> D (IN REF. 3).
P -> S (IN REF. 3).
Q -> H (IN REF. 3).
Q -> H (IN REF. 3).
STTQFHAGMR -> ALSITTSQHSAPTAPELITLS
SLG (IN ISOFORM 6).
K -> KO.
/FTID=VAR.010926.
E -> Q (IN REF. 3).
M -> I (IN REF. 3).
E -> G (IN REF. 3).
R -> S (IN REF. 3).
N -> S (IN REF. 3).
H -> N (IN REF. 3).
K -> N (IN REF. 3).
OKH -> PKP (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
N -> D (IN REF. 3).
V -> E (IN REF. 3).
R -> P (IN REF. 1 AND 2).
N -> T (IN REF. 3).
V -> G (IN REF. 3).
MISSING (IN REF. 5).
A -> P (IN REF. 3).
I -> V (IN REF. 3).
V -> D (IN REF. 5).
EI -> HE (IN REF. 7).
E -> V (IN REF. 3).
L -> R (IN REF. 3).
I -> N (IN REF. 8).
E -> D (IN REF. 3).
P -> S (IN REF. 3).
Q -> H (IN REF. 3).
Q -> H (IN REF. 3).
STTQFHAGMR -> ALSITTSQHSAPTAPELITLS
SLG (IN ISOFORM 6).
K -> KO.
/FTID=VAR.010926.
E -> Q (IN REF. 3).
M -> I (IN REF. 3).
E -> G (IN REF. 3).
R -> S (IN REF. 3).
N -> S (IN REF. 3).
H -> N (IN REF. 3).
K -> N (IN REF. 3).
OKH -> PKP (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
N -> D (IN REF. 3).
V -> E (IN REF. 3).
R -> P (IN REF. 1 AND 2).
N -> T (IN REF. 3).
V -> G (IN REF. 3).
MISSING (IN REF. 5).
A -> P (IN REF. 3).
I -> V (IN REF. 3).
V -> D (IN REF. 5).
EI -> HE (IN REF. 7).
E -> V (IN REF. 3).
L -> R (IN REF. 3).
I -> N (IN REF. 8).
E -> D (IN REF. 3).
P -> S (IN REF. 3).
Q -> H (IN REF. 3).
Q -> H (IN REF. 3).
STTQFHAGMR -> ALSITTSQHSAPTAPELITLS
SLG (IN ISOFORM 6).
K -> KO.
/FTID=VAR.010926.
E -> Q (IN REF. 3).
M -> I (IN REF. 3).
E -> G (IN REF. 3).
R -> S (IN REF. 3).
N -> S (IN REF. 3).
H -> N (IN REF. 3).
K -> N (IN REF. 3).
OKH -> PKP (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
N -> D (IN REF. 3).
V -> E (IN REF. 3).
R -> P (IN REF. 1 AND 2).
N -> T (IN REF. 3).
V -> G (IN REF. 3).
MISSING (IN REF. 5).
A -> P (IN REF. 3).
I -> V (IN REF. 3).
V -> D (IN REF. 5).
EI -> HE (IN REF. 7).
E -> V (IN REF. 3).
L -> R (IN REF. 3).
I -> N (IN REF. 8).
E -> D (IN REF. 3).
P -> S (IN REF. 3).
Q -> H (IN REF. 3).
Q -> H (IN REF. 3).
STTQFHAGMR -> ALSITTSQHSAPTAPELITLS
SLG (IN ISOFORM 6).
K -> KO.
/FTID=VAR.010926.
E -> Q (IN REF. 3).
M -> I (IN REF. 3).
E -> G (IN REF. 3).
R -> S (IN REF. 3).
N -> S (IN REF. 3).
H -> N (IN REF. 3).
K -> N (IN REF. 3).
OKH -> PKP (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
N -> D (IN REF. 3).
V -> E (IN REF. 3).
R -> P (IN REF. 1 AND 2).
N -> T (IN REF. 3).
V -> G (IN REF. 3).
MISSING (IN REF. 5).
A -> P (IN REF. 3).
I -> V (IN REF. 3).
V -> D (IN REF. 5).
EI -> HE (IN REF. 7).
E -> V (IN REF. 3).
L -> R (IN REF. 3).
I -> N (IN REF. 8).
E -> D (IN REF. 3).
P -> S (IN REF. 3).
Q -> H (IN REF. 3).
Q -> H (IN REF. 3).
STTQFHAGMR -> ALSITTSQHSAPTAPELITLS
SLG (IN ISOFORM 6).
K -> KO.
/FTID=VAR.010926.
E -> Q (IN REF. 3).
M -> I (IN REF. 3).
E -> G (IN REF. 3).
R -> S (IN REF. 3).
N -> S (IN REF. 3).
H -> N (IN REF. 3).
K -> N (IN REF. 3).
OKH -> PKP (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
N -> D (IN REF. 3).
V -> E (IN REF. 3).
R -> P (IN REF. 1 AND 2).
N -> T (IN REF. 3).
V -> G (IN REF. 3).
MISSING (IN REF. 5).
A -> P (IN REF. 3).
I -> V (IN REF. 3).
V -> D (IN REF. 5).
EI -> HE (IN REF. 7).
E -> V (IN REF. 3).
L -> R (IN REF. 3).
I -> N (IN REF. 8).
E -> D (IN REF. 3).
P -> S (IN REF. 3).
Q -> H (IN REF. 3).
Q -> H (IN REF. 3).
STTQFHAGMR -> ALSITTSQHSAPTAPELITLS
SLG (IN ISOFORM 6).
K -> KO

DR PRINTS: PRO0380; KINESINHEAVY.
 DR SMART; SM00129; KISC; 1.
 DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
 DR PROSITE; PS50067; KINESIN MOTOR DOMAIN2; 1.
 KW Motor protein: Microtubules; ATP-binding; Coiled coil; Mitosis;
 Phosphorylation.
 FT DOMAIN 9 356 MECHANOCHEMICAL (MOTOR) (BY SIMILARITY).
 FT DOMAIN 357 479 COILED COIL (POTENTIAL).
 FT DOMAIN 688 729 COILED COIL (POTENTIAL).
 FT DOMAIN 868 895 COILED COIL (POTENTIAL).
 FT DOMAIN 1033 1060 COILED COIL (POTENTIAL).
 FT NP_BIND 98 105 ATP (BY SIMILARITY).
 FT MOD_RES 930 930 PHOSPHORYLATION (BY CDC2) (BY
 SIMILARITY).
 SQ SEQUENCE 1060 AA; 119332 MW; 201E7EB6962B3492 CRC64;

Query Match 3.18; Score 164; DB 1; Length 1060;
 Best Local Similarity 20.38; Pred. No. 0.054;
 Matches 214; Conservative 155; Mismatches 383; Indels 302; Gaps 52;

QY 24 TDDLEFVWNLNREVNINICEKEVEODARGII---HMLIKGSESCNLF---LKSLEK 76
 114 SDEEFW-----EODPLAGIIPRLHQIFKEKSEKTEFSYKVSLEIY 157
 QY 77 NYPLEFDLN---GQSL---FHQTSSEG---DLDDLAQDKD-LYH-----T 111
 158 NEELFPLSLSPSPVGRLOQFDDPRNKRGVYIKGLEISVHNKDEYHILDEGARRKTA 217
 QY 112 PSFLNFPYPLGED--IDILFNLKSTFEPRVLMKRDQHHHREQLTNGLLQALOSPCLIEG 169
 218 STLMNLYSSRSVSFVTLHMKKTEYDVG-----EELVYKIGKMLVD---LAG 261
 QY 170 ESKGKSTLQRIAMLMGSGCKALTKFKVFPL-----RLSRAQGLFETLDOLDIP 224
 262 SEVIGSGAVDKARAGNINOSLLTGLRYTALVETPHIIPRESKTRPIIDSL----- 317
 QY 225 GTIRKOTFAMM---LKRQRYLFLDGYNEKP-ONCEPI-----EALIKENHFKN 273
 318 GGRKTSIIATVSPASINIEEV-STLDYANRAKSIMNKEVNOQLTKKLKLIKE----- 370
 QY 274 MVIYTTTECLRIHIOFGALTAEVG-----DMTEDSAQALLREVLKELAEDELLQIQK 327
 371 -----YTEELERLKRRLAAREKNVYLSENEDLOGKVLSEEMITEYETKITAAMEE 425
 QY 328 SRCLRNIMKTPLEVVITCA--IQMGSEFHSHTQTLFHTFYLLIQKNHKKRGAASD 385
 426 LKISISLFPADNKKKELECTTIIQCKEKL-EETQNH-----QESKEQ---LAQES 472
 QY 386 FIRSLDHCGDLEGVFSHKFDELDQVSSVNEVDLLTGLCKYTAQRPKPYKRF-FHK 444
 473 FVVSARETEKRLHGT-ANKL-----LSTVRETRDVSGSLHKKLDRKKAVDQHNQVHE 525
 445 SPOEYTAGRRLSLTSHEPEYTKNGYLOKAV-SISDTTSYSSLRTGSSSEATR 503
 526 NPAE-OMDRRFVSIQRTVDYSV-KOOGMLDPTNSIDDLGASSSRSLATAS----- 576
 QY 504 AVMKHLAAYYOHCLGLSLAKRPLRROESIOSVKNTEDEILKAININSFVCGIHLYO 563
 577 AVAKSPASV-----QETVTPQVSHSVSEIILKQETLS----- 607
 QY 564 ESTSKSALSOEFAPFOGSLYINSNIP--DYLPDFEHLPMCASALDEI--KLPYGG 619
 608 -SOAKEDLOOLMAAHRTGLEALRSDLLPVATVLDLNSHSLCLONFLIVADKIDSHKE 666
 QY 620 AMASWEKAEDNGIMEAPEYIIPSAVSLFFENKOEKERTIE----- 663
 667 DMNSF--FTEHSRSLHKLRLDSSSLSSIQSESESLKEDATIAQSMHSEGVNLISSLO 724
 QY 664 -----VTLRDS-----KLNK-----ODITYLG-----KISSATSLR 691
 725 QUNLGMETQOQSGPLSKGKRIQKSVGSLQDDLDLVSSAIEICISHHKKLAQSDOVA 764

QY 692 LQIKRQAGVAGSLSLVSTCKNIYSLMEVNSPLTIDE-----RHITSVNLKLTLSH 744
 785 VEIROLAG--SNMSTLEESKQCEKLTSSINTISQSQWCESAGQKIDSVLEQYCYLH 842
 QY 745 DLON--QRLPGLGTDLSGN-----LKNLTKLIMDINKNEEDAIKL 783
 843 SSRKHLONLKRGVEDSCGSSVEITRVNAQROAEKALTSLEQVRDQEMVGEORLEL 902
 QY 784 ABEKLN-LKKMCFEHLTHL-SDIGEG-----MDYIKSL--SSEPCDL-----EE 824
 903 QEOVQSGLNKVVHSLYEELRLNDVPTGTTPQRDYATPSSLVYKTPRDVLLQFRQOOOEY 962
 QY 825 IQLVSCCLSANAVK-----ILAQNLHNLVKSITLDSENVLEKDG----- 864
 963 LESISSVYS-EAVEPVEDQSLDEPPVAVNDVSIERSCIDLSMTCKQKGGIRFPOKK 1021
 QY 865 -----NEALHELIDRMNVLEQ-----LRLALMP 887
 1022 ALRKEKENRGNTTLERSKIMDEVQALTKSKLP 1055

RESULT 11

SPCA_HUMAN STANDARD; PRT; 2418 AA.
 ID P02349; Q15514;

DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE SPECTRIN ALPHA CHAIN, ERYTHROCYTE.
 GN SP7AI OR SP7A.

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 NCBI_TaxID:9606;

RN [1]
 RP MEDLINE=90170949; PubMed=1689726;
 RX Sahr K.E., Laurila P., Kotula L.,
 RA Linenbach A.J., Winkelman J.C., Speicher D.W., Marchesi V.T.,
 RA Curtis P.J., Forget B.G.;
 RT "The complete cDNA and polypeptide sequences of human erythroid
 alpha-spectrin.";
 RT J. Biol. Chem. 265:4434-4443(1990).

RN [2]
 RP MEDLINE=90170949; PubMed=1689726;
 RX Sahr K.E., Tope T., Scarpa A., Laughinghouse K., Marchesi S.L.,
 RA Agre P., Linenbach A.J., Marchesi V.T., Forget B.G.;
 RT "Sequence and exon-intron organization of the DNA encoding the alpha
 I domain of human spectrin. Application to the study of mutations
 causing hereditary elliptocytosis.";
 RT J. Clin. Invest. 84:1243-1252(1989).

RN [3]
 RP MEDLINE=84087888; PubMed=6654896;
 RX Speicher D.W., Davis G., Marchesi V.T.;
 RT "Structure of human erythrocyte spectrin. II. The sequence of the
 alpha-I domain.";
 RT J. Biol. Chem. 258:14938-14947(1983).

RN [4]
 RP MEDLINE=84087887; PubMed=6654895;
 RX Speicher D.W., Davis G., Yurchenco P.D., Marchesi V.T.;
 RT "Structure of human erythrocyte spectrin. I. Isolation of the alpha-I
 domain and its cytosolic bromide peptides.";
 RT J. Biol. Chem. 258:14931-14937(1983).

RN [5]
 RP MEDLINE=86205962; PubMed=3458204;
 RX Linenbach A.J., Speicher D.W., Marchesi V.T., Forget B.G.;
 RT "Cloning of a portion of the chromosomal gene for human erythrocyte
 alpha-spectrin by using a synthetic gene fragment.";
 RT Proc. Natl. Acad. Sci. U.S.A. 83:2397-2401(1986).

RN [6]
 RP PARTIAL SEQUENCE.
 RX MEDLINE-84295638; PubMed-6472478;
 RA Speicher D.W., Marchesi V.T.;
 RT "Erythrocyte spectrin is comprised of many homologous triple helical
 segments.";
 RL Nature 311:177-180(1984).
 RN [7]
 RP IDENTIFICATION OF PROBABLE FRAMESHIFT IN 2407-2418.
 RA Gibson T.J.;
 RL Unpublished observations (MAR-1995).
 RN [8]
 RP REVIEW ON VARIANTS.
 RX MEDLINE-97001215; PubMed-8844207;
 RA Millet P., Alloisio N., Morle L., Delaunay J.;
 RT "Spectrin mutations in hereditary elliptocytosis and hereditary
 spherocytosis.";
 RL Hum. Mutat. 8:97-107(1996).
 RN [9]
 RP VARIANT SER-24.
 RX MEDLINE-94289716; PubMed-8018926;
 RA Parquet N., Devaux I., Boulanger L., Galand C., Boivin P.,
 RL Lecomte M.-C., Dhermy D., Garbarg M.;
 RT "Identification of three novel spectrin alpha I/74 mutations in
 hereditary elliptocytosis: further support for a triple-stranded
 folding unit model of the spectrin heterodimer contact site.";
 RL Blood 84:303-308(1994).
 RN [10]
 RP VARIANTS CYS-28; HIS-28; LEU-28 AND SER-28.
 RX MEDLINE-91358728; PubMed-1679439;
 RA Coetzer T.L., Sahr K., Prechal J., Blacklock H., Peterson L., Koler R.,
 Doyle J., Manaster J., Palek J.;
 RT "Four different mutations in codon 28 of alpha spectrin are
 associated with structurally and functionally abnormal spectrin alpha
 I/74 in hereditary elliptocytosis.";
 RL J. Clin. Invest. 88:743-749(1991).
 RN [11]
 RP VARIANTS SER-28 AND ARG-48.
 RX MEDLINE-91346849; PubMed-1878597;
 RA Floyd P.B., Gallagher P.G., Valentini L.A., Davis M., Marchesi S.L.,
 Forget B.G.;
 RT "Heterogeneity of the molecular basis of hereditary
 pyropoikilocytosis and hereditary elliptocytosis associated with
 increased levels of the spectrin alpha I/74-kilodalton tryptic
 peptide.";
 RL Blood 78:1364-1372(1991).
 RN [12]
 RP VARIANT SER-45.
 RX MEDLINE-89323468; PubMed-2568862;
 RA Lecomte M.-C., Garbarg M., Grandchamp B., Feo C., Gautero H.,
 Devaux I., Bounnier O., Galand C., D'Auriol L., Gilbert F.,
 Sahr K.E., Forget B.G., Boivin P., Dhermy D.;
 RT "Sp alpha I/78: a mutation of the alpha I spectrin domain in a white
 kindred with HE and HPP phenotypes.";
 RL Blood 74:1126-1133(1989).
 RN [13]
 RP VARIANT PRO-207.
 RX MEDLINE-92176375; PubMed-1541680;
 RA Gallagher P.G., Tse W.T., Coetzer T., Lecomte M.-C., Garbarg M.,
 Zarkowsky H.S., Baruchel A., Ballas S.K., Dhermy D., Palek J.,
 Forget B.G.;
 RT "A common type of the spectrin alpha I 46-50a-kD peptide abnormality
 in hereditary elliptocytosis and pyropoikilocytosis is associated
 with a mutation distant from the proteolytic cleavage site. Evidence
 for the functional importance of the triple helical model of
 spectrin.";
 RL J. Clin. Invest. 89:892-898(1992).
 RN [14]
 RP VARIANT VAL-1857.
 RX MEDLINE-93253053; PubMed-8486776;
 RA Wilmotte R., Marechal J., Morle L., Baklouti F., Philippe N.,
 Kastally R., Kotula L., Delaunay J., Alloisio N.;
 RT "Low expression allele alpha Ilevy of red cell spectrin is associated

RT with mutations in exon 40 (alpha V/41 polymorphism) and intron 45 and
 RT with partial skipping of exon 46.";
 RL J. Clin. Invest. 91:2091-2096(1993).
 RN [15]
 RP VARIANT BARCELONA.
 RX MEDLINE-93372367; PubMed-8364215;
 RA dalla Venezia N., Alloisio N., Forissier A., Denoroy L., Aymerich M.,
 Vives-Corrons J.L., Besalduch J., Besson I., Delaunay J.;
 RT "Elliptopikilocytosis associated with the alpha 469 His-->Pro
 mutation in spectrin Barcelona (alpha I/50-46b).";
 RL Blood 82:1661-1665(1993).
 RN [16]
 RP VARIANT CAGLIARI.
 RX MEDLINE-94043025; PubMed-8226774;
 RA Sahr K.E., Coetzer T.L., Moy L.S., Derick L.H., Chishti A.H.,
 Jarolim P., Lorenzo F., Miraglia del Giudice E., Iolascon A.,
 Gallanelli R.;
 RT "Spectrin Cagliari: an Ala-->Gly substitution in helix 1 of beta
 spectrin repeat 17 that severely disrupts the structure and self-
 association of the erythrocyte spectrin heterodimer.";
 RL J. Biol. Chem. 268:22656-22662(1993).
 RN [17]
 RP VARIANTS CULOZ AND LYON.
 RX MEDLINE-90347052; PubMed-2384601;
 RA Morle L., Roux A.-F., Alloisio N., Pothier B., Starck J., Denoroy J.,
 Morle F., Rudigoz R.-C., Forget B.G., Delaunay J., Godet J.;
 RT "Two elliptocytogenic alpha I/74 variants of the spectrin alpha I
 domain. Spectrin Culoz (G67-->G77; alpha I 40 Gly-->Val) and
 spectrin Lyon (C77-->T77; alpha I 43 Leu-->Phe).";
 RL J. Clin. Invest. 86:548-554(1990).
 RN [18]
 RP VARIANT JENDOURA.
 RX MEDLINE-92345619; PubMed-1638030;
 RA Alloisio N., Wilmotte R., Morle L., Baklouti F., Marechal J.,
 Ducloux M.-T., Denoroy L., Feo C., Forget B.G., Kastally R.,
 Delaunay J.;
 RT "Spectrin Jendouba: an alpha II/31 spectrin variant that is
 associated with elliptocytosis and carries a mutation distant from
 the dimer self-association site.";
 RL Blood 80:809-815(1992).
 RN [19]
 RP VARIANT TUNIS.
 RX MEDLINE-89323436; PubMed-2568861;
 RA Morle L., Morle F., Roux A.F., Godet J., Forget B.G., Denoroy L.,
 Garbarg M., Dhermy D., Kastally R., Delaunay J.;
 RT "Spectrin Tunis (Sp alpha I/78), an elliptocytogenic variant, is due
 to the C6G-->T6G codon change (Arg-->Trp) at position 35 of the
 alpha I domain.";
 RL Blood 74:828-832(1989).
 RN [20]
 RP VARIANT GENOVA.
 RX MEDLINE-94250920; PubMed-8193371;
 RA Perrotta S., del Giudice E.M., Alloisio N., Sclarratta G., Pinto L.,
 Delaunay J., Cuttillo S., Iolascon A.;
 RT "Mild elliptocytosis associated with the alpha 34 Arg-->Trp mutation
 in spectrin Genova (alpha I/74).";
 RL Blood 83:3346-3349(1994).
 RN [21]
 RP VARIANT ANASTASIA.
 RX MEDLINE-95290423; PubMed-7772539;
 RA Perrotta S., Iolascon A., de Angelis F., Pagano L., Colonna G.,
 Cuttillo S., del Giudice E.M.;
 RT "Spectrin Anastasia (alpha I/78): a new spectrin variant (alpha 45
 Arg-->Thr) with moderate elliptocytogenic potential.";
 RL Br. J. Haematol. 89:933-936(1995).
 CC -I- FUNCTION: SPECTRIN IS THE MAJOR CONSTITUENT OF THE CYTOSKELETAL
 NETWORK UNDERLYING THE ERYTHROCYTE PLASMA MEMBRANE. IT ASSOCIATES
 WITH BAND 4.1 AND ACTIN TO FORM THE CYTOSKELETAL SUPERSTRUCTURE OF
 THE ERYTHROCYTE PLASMA MEMBRANE.
 CC -I- SUBUNIT: COMPOSED OF NONHOMOLOGOUS CHAINS, ALPHA AND BETA, WHICH
 AGGREGATE SIDE-TO-SIDE IN AN ANTIPARALLEL FASHION TO FORM DIMERS,
 TETRAMERS, AND HIGHER POLYMERS.
 CC -I- DOMAIN: SPECTRIN-LIKE PROTEINS HAVE FIVE DOMAINS:

CC (1) N-TERMINAL DOMAIN (N),
 CC (2) DOMAIN BETWEEN THE N-TERMINAL AND MIDDLE DOMAIN (NM),
 CC (3) MIDDLE DOMAIN (M),
 CC (4) DOMAIN BETWEEN THE MIDDLE AND C-TERMINAL DOMAIN (MC),

Query Match 2.9%; Score 152.5; DB 1; Length 2416;

Best Local Similarity 21.3%; Pred. No. 0.76; Matches 198; Conservative 128; Mismatches 319; Indels 285; Gaps 52;

176 STLLRIAMLSGSKCAL--TKFKV-----FFLRSLAAGLPE---TLGDO-LLDI 223
 1472 ATRLQVLDNRKALQDLDERKIDYANKQFYRDLLEWISMLPTACDESKDA 1531
 224 PGTRK---QTFMAMLKIRQVFLLDGYNEFKPQNCPEIALIKE----- 267
 1532 TNIQRKYLKQTFANHEVDGSEQVGVINIGNSLIERSCGNEAMKEQLEKEMDH 1591
 268 ---NIRFKNMVIVTTTECLRIHIFOGALTA-----VGMTESA-----QAL 308
 1592 LERTNKGKRLNASHKQOQRENTSIRDFEMLSEETLLAKMDARLASAGNLKKHOLL 1651
 309 IREVLKE-----LAEGILL-----QIQSRCLRLMKTPFVVITCAIQMGES 353
 1652 ERMMLAREDAKDLNTLAEGLSSGTFNVDQYKK--DNVKNRFL-----NVQELAA 1703
 354 FHSHTQ--TLFHTFYDLLIQN--KKHKVYASDFINSLDHGDL-----ALEG-VPS 403
 1704 HHKRLKVAALFOFQODLDEESIMEEKLIRVSSQDYGRDQVQNLKKHKLREGELVA 1763
 404 HKRDFLOVSSVNEVDVLLTGILCTYQRFKPKFKFHSQETAGRLSLTSH 463
 1764 H--EPALQVLDL-----AEKLDKAAVQOELO-----LRLAQVEHW 1801
 464 P--EYVKGNG-----YLOKMSISD-----ITSTYSLRLTYCGSSVEATRA-VMKHL 509
 1802 KKLKELAKAGIKLEESLEVIQFMQNAEBEAWNEKNALAVRDCDPTLAATQSLMKHE 1861
 510 A-----AVY-----QHCGLLGLSTAKRPLMRQESLSQVKNTE-----QELIKAININSE 554
 1862 ALENDRVAHETRYQNVCAQGEDILNKVLEHSQNKESIKLEALNKETPSIAKAIAMKL 1921
 555 VEGGILHYESTSKALSQEFAPFOGKSIYI--NSGNIPDYLFDF----- 599
 1922 QLEDYDAFOQFMKADY--VEAMWADKETSLENTNGSAD--LGDFTLLAKQDPTLASIQ 1977
 600 ---EHLRNCASALDFIKL-----DFYGGAMASWEKAEDTGGIHMEAPETY 643
 1978 SFGQERLPEITDQK--KLISAGHNSKAIIEERYAALLKRWEDLLE-ASAHHQKILEKQ 2034
 644 IP-SRAVSLFF-----NM--KOEFTLE---VILRDFSKLNKODITYLKI--- 683
 2035 LPLQKEDLEFEFAHNASALNNCEKMEENLSEPVHCVSINEIROLOKHEDFLASARA 2094
 684 ---FSSATSLRLQIKRQAGVAGS-----LSLVSTCKNYSIMAVE-ASPLTIEDERHIT 733
 2095 QADFKLLELDQOIK-ALGVPPSPYTWLYVEYERTWKHLSDIIEEREDLOKEEAROVK 2153
 734 SV-----TNKTLSDIHLQONR--LPGLDLSLGNLKNLTKLINDIKMNEDEAIKLA 784
 2154 NFMQOEFGQNASTFLQWILETRAYFLDGLSLKETGTL--SOLLEANKRKQKQIKQAMKR- 2210
 785 EGLKLNKMLCPHLTHLSDIGEGMDYIVKSLSEPCDLEIQLVSCCLSANAVKIIAONL 844
 2211 ---OLTKITVLDGD-----NLEDAITLD----- 2229
 845 HNLVKLSIIDLSENYLEKNGNEALHELIDRM--NVLEQLTALMLPWCQVQ-----S 895
 2230 ---IKSTIGLQOQW-----DQVQLGRLMOHNLQOIQAK-----DIKGVSEETLKE 2274
 896 LSLKLNHEEVPOLVTKLGNRLTDEIR 925
 2275 FSTYVKNHDE-----NLGRLTIHKEFR 2296

RESULT 12
 YMW6_YEAST
 ID YMW6_YEAST STANDARD; PRT: 1277 AA.
 AC 004264; 004780;

DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HYPOTHEICAL 147.0 KDA PROTEIN IN ABF2-CHL12 INTERGENIC REGION.
 GN YMR076C OR YMR582.01C OR YMR916.15C.
 OS Saccharomyces cerevisiae (baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RX [1]
 RP SEQUENCE OF 1-682 FROM N.A.
 RC STRAIN=S288C / AB972;
 RA Gentles S., Bowman S., Barrell B.G., Rajandream M.A.;
 RL Submitted (May-1995) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 659-1227 FROM N.A.
 RC STRAIN=S288C / AB972;
 RA Pearson D., Bowman S., Barrell B.G., Rajandream M.A., Walsh S.V.;
 RL Submitted (Apr-1995) to the EMBL/Genbank/DBJ databases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: 249259; CAA89222.1;
 CC DR EMBL: 248952; CAA88801.1;
 CC DR SGD: S0004681; YMR076C.
 KW Hypothetical protein.
 SQ SEQUENCE 1277 AA; 147040 MW; 9DF40A5274FD9623 CRC64;

Query Match 2.8%; Score 150; DB 1; Length 1277;
 Best Local Similarity 18.8%; Pred. No. 0.44; Matches 183; Conservative 146; Mismatches 354; Indels 286; Gaps 41;

270 RKNMVIYV---TTTECLRIHIFOGALTAEGVGTEDSA-----QALIREVLIR 315
 10 KENSPIISTSDQISINELDLRK--ALHHEFLASIDQONTDLTGDXRYDALVSKLKL 66
 316 ELAEGILLIQSRCLRNLMK-----PLF--VITCAIQMGESFHSHTQTT 361
 67 HKDVG--IRAFTRACCLSDILRLYAPDPAPYTDQLDIFKLVLSQFEBOLDQENGHIQOT 124
 362 ---LFTFYDLLIQNK-----HKHKGVAASDFR 388
 125 YLTKLLEKRSIYLLADLPSSNNLILFHFID---PNKSPRALFNVIGIIGEVIS 180
 389 SLDHCGDLALLEGVSHKDFE---ELQDVSSVNEVDVLLTGL-LCKYTAQRFK--PKY-- 439
 181 EFDVSPLLEVLRILFNKFLYNPNEIPEGNVSDGCVESLIDCDYYSRMSRHLTKYYS 240
 440 KFFKHSFOETAGRLSLTSH-----PEVYTGNGYLOKMS----- 479
 241 EITHETNDNNSRLTLVYVVKHKLVLRLMEYVPELDINAVIGIYHIELSENELEFKREAT 300
 480 ---ISDTSYSSSLRYTCSS-----VEATRAYVKHLAAVYQHGCLLG 520
 301 KLIGQILLTISDINFTVSTSDTFKAMISKIADISPDVAVEMTESIPQIATBEDISKELN 360
 521 LSIARK-----PLMRQESLSQVKNTEQELIKAININSEFEGIHLYQESTK----- 568
 361 QALAKFTISDPRVTRTSVMFKKVPYELMKWNITKAIYTSILHARKHKEVRELCIN 420
 569 ---SALSQEFAPFOGKSLYINSGNIPDYLF--FFHLRNCASALDFIKLDYGGAM 621


```

Db 684 NVLDSKSVVEYIOT-LAYMELQOKAEFSQOKHOKELIENMCKLT-----SOLAGOVE 733
OY 88 SLFH--QTSBGDDDLAODLKDLYHTPSFLNFPYLGEDIDIFN-----LKSTFEPLV 139
Db 734 DLEHKQILSLNEMDMRCQODLHAEEESRLDLKSKDASLVNEDHQRBLLAFDQOPAM 793
OY 140 WRKQDHHHRYEQLTLNGLLOALOSPCTIEGESKSKSTLIO-RIAMLWSSGCKALTKFK 198
Db 794 -----HHSFANIIGEOGSMPSERSECRLEADQSPKNSAILQNNVDSL----- 835
OY 199 FVFFELSLRAGQGLFELIDCDLIDPTIKQOFMAMLLKRLQRVFLDLGYNFKPQNC 258
Db 836 -BESTSQOMASDLQOCCELVOIKGEIEN-----LMKAEQM-----HOSFAETS 882
OY 259 PEIEALIKENHFRKNVIVTTTECLNHRQFALTAEVGDMTEDSKOALIREV-LIKEL 317
Db 883 QRSKIQEDISAQN-VVAFETSALENKEKELELDLNDK-----EFQOAIQELKSNHL 936
OY 318 AEGLLIQOKSRCLNLMKTPLEVVITCAIQMGSEFHSHTQ--TLEFHTFYDL----- 369
Db 937 LEDSLLEQLLSEPLEKKEKMSIIS--LTKREIEELTQENGTLEKINASINQEKMN 992
OY 370 LIOKKN-----HKHGVV-ASDFIRS-----LDHOGDL--ALEGVSHKPELQDVS 414
Db 993 LIOKSSFPANYIDEREKSISELSQYKOEKLLIQCEETGMAVEDL-SQYK-AAQEK 1050
OY 415 SVNEDVLLTGLCKYTAORFKRYEFHKSPOEY-----TAGRRL 455
Db 1051 SKLECLNCTSLCENKKNKELEQLEKAFKEHQBFLTKLAFABERNQNLMELEYQOAL 1110
OY 456 SLLTSHREPEVYKNGYLOKMWISD---ITSTYSLLRITCGSSVEVTRAVMKHLAA 511
Db 1111 RSEMTQNNNSKSEAGLKEIMTLKEONKMOKEVNDLQEN-----EQLMKVMK---- 1161
OY 512 VYQHGCLLGLSIKAPLPMROESLOS--VKNTTQOEILKAININSFVCGIHLQOES-TSK 568
Db 1162 -THHEC-----QNESEPIRNSVVERESERNQCFKQOMDELEVGEISLSDY 1206
OY 569 SALSQFEAFQOKSLYXNSGNTPDVLFDFEHLPNCAS-ALDEIKLDFGGAMASWEKA 627
Db 1207 NAQVLEALRNKELKLOES-----EKEKECLOHELOTIRDELTSNLQDMQ-- 1254
OY 628 AEDTQGIHNE--APEYIIPS-RAVSLFFMWKQEF--TLEVTLRDPSKLMKODITYIGKI 663
Db 1255 SQPISGLKQCEIDAEERYISGPHELSTSONDMAHLOCSLQTTWKNLKELEICEIIQAEK 1314
OY 684 FSSATSLRLQIKRCAGVAGSLVLTCKNIYSLNWEASPL-----TIED----- 728
Db 1315 YELVTELENDRSRCITATRKMAEEVGKLNVEKILINDSGELGELVEDLPGEGFQOPN 1374
OY 729 ERHITSVTNL-----KTLSTHDLQNRPLPGLTDSLGNLKNLTKLMD----- 771
Db 1375 EQHPVSLAPIDESNSYEHLTLSDKEVQOMHFAEQEKPLSDQSEHKILHDQCOMSKMSE 1434
OY 772 -----NIKMNEDAIK-----LABEL-KNLKMKCLFHLTHSDIGEGM 808
Db 1435 LQTYVDSLAEMLVLTSTNLNPFQDVLKEMQGLIEBLVPSLSSCCVPDSSSLSDSS 1494
OY 809 DY-----IYKSLSEPDDEIQLVSCCLSANAVK-----TLAOLHN--- 846
Db 1495 FYRALLEQTDMSLNLNBEAVSANCOSYDEVFCSSIQTYVDSIKAEMLVLTSTNLNRFQ 1554
OY 847 -LVKLSILDSENYL-----EKDQNEALHEDIDRMNVLEOULTALMLP 887
Db 1555 DLYKEMQGLIEBLVPSLSSCCVPDSSSLSDSSFYRALLBQTDGMSLNLNLEGVVA 1614
OY 888 WGDVQSGSLSLKH-----LEVPOLVKLGLKMMKRLTDTETIRLGAFFGKNPLKNFQ 943
Db 1615 NQOSVBEVFCSSLOEBENLRKRETPSAPAKVE-----ELSEICEYR----- 1660
OY 944 LAGNRVSDGMLAFMGVFNULQV-----PDSFKFELPDPAVLVKLSOVLSKLFLQ 998
Db 1661 KLEEKHESQIMKNKEI-QELBOLLSSEROELDLCKROYLSEN-----EQMOOKLTSVLE 1716

```

```

OY 999 EARLVGWQFDDDDLSV 1014
Db 1717 ESKLAEKKQTEQLSL 1732

RESULT 14
BEM2_YEAST
ID BEM2_YEAST STANDARD; PRT; 2167 AA.
AC P39960;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GTPASE ACTIVATING PROTEIN BEM2/1P12.
OS BEM2 OR 1P12 OR SUP9 OR YER155C.
OC Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=528BC;
RX MEDLINE=95050996; PubMed=7962097;
RA Kim Y., Francisco L., Chen G., Marcotte E., Chan C.S.;
RT Control of cellular morphogenesis by the 1p12/Bem2 GTPase-activating
RL protein: possible role of protein phosphorylation.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95050997; PubMed=7962098;
RA Peterson J., Zheng Y., Bender L., Myers A., Cerione R., Bender A.;
RT Interactions between the bud emergence proteins Bem1p and Bem2p and
RL Rho-type GTPases in yeast.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=528BC / AB972;
RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,
RA Aviles E., Berno A., Brennan T., Carpenter J., Chen E., Cherry J.M.,
RA Chung E., Duncan M., Guzman E., Hartzell G., Hunkle-Smith S.,
RA Hyman R., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,
RA Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,
RA Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,
RA Taylor P., Wei Y., Yellon M., Botstein D., Davis R.W.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=94123964; PubMed=8293973;
RA Chan C.S., Botstein D.;
RT Isolation and characterization of chromosome-gain and increase-in-
RT ploidy mutants in yeast.
RL Genetics 135:677-691(1993).
CC -1- FUNCTION: GTPASE-ACTIVATING PROTEIN (GAP) FOR RHO1 AND RHO2.
CC INVOLVED IN THE CONTROL OF CELLULAR MORPHOGENESIS. REQUIRED FOR
CC PROPER BUD SITE SELECTION AND BUD EMERGENCE.
CC -1- SIMILARITY: CONTAINS 1 RASGEF DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 GAP DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: Z35159; CA84524.1; -
CC DR EMBL: L33832; AAA57132.1; -
CC DR EMBL: U18917; AAB64682.1; -
CC DR SGD: S0000957; BEM2.
CC DR InterPro: IPR001849; PH.
CC DR InterPro: IPR000651; RASGEFN.

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 25, 2002, 10:55:45 ; Search time 35.13 Seconds
(without alignments)
4263.674 Million cell updates/sec

Title: US-09-697-089-2

Perfect score: 5281

Sequence: 1 MNEFKDNRRLIQRMGMTVI.....WQFDDDLVITGAFKLVT 1024

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

rched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_17:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2244	42.5	439	4 Q9NPP4	Q9NPP4 homo sapien
2	296	5.6	953	4 Q9Y239	Q9Y239 homo sapien
3	266	5.0	1062	4 Q9H666	Q9H666 homo sapien
4	264	5.0	1062	4 Q9NX02	Q9NX02 homo sapien
5	264	5.0	1062	4 Q9BVN5	Q9BVN5 homo sapien
6	261.5	5.0	1033	4 Q9HAV9	Q9HAV9 homo sapien
7	242	4.6	1111	11 Q9R1M5	Q9R1M5 mus musculu
8	241	4.6	1111	11 Q9JLR2	Q9JLR2 mus musculu
9	234.5	4.4	1040	4 Q9HC29	Q9HC29 homo sapien
10	217	4.1	1429	4 Q9HAV8	Q9HAV8 homo sapien
11	212	4.0	1192	4 Q9UFT4	Q9UFT4 homo sapien
12	212	4.0	1429	4 Q9Y2E0	Q9Y2E0 homo sapien
13	207.5	3.9	1473	4 Q9C000	Q9C000 homo sapien
14	207.5	3.9	1399	4 Q9BZ29	Q9BZ29 homo sapien
15	207.5	3.9	1443	4 Q9BZ29	Q9BZ29 homo sapien
16	201	3.8	1031	10 Q9FR29	Q9FR29 arabidopsis
17	196.5	3.8	1590	10 Q9FTB5	Q9FTB5 arabidopsis
18	187	3.5	1405	10 Q50052	Q50052 arabidopsis
19	183	3.5	748	11 Q9NMW0	Q9NMW0 mus musculu

20	183	3.5	1187	10 Q9F114	Q9F114 arabidopsis
21	181	3.4	490	4 Q9H5N5	Q9H5N5 homo sapien
22	180.5	3.4	1054	7 Q78109	Q78109 mus musculu
23	180.5	3.4	1135	7 Q9TPE1	Q9TPE1 mus musculu
24	178.5	3.4	1031	10 Q9CAK0	Q9CAK0 arabidopsis
25	176.5	3.3	1108	10 Q9ARB2	Q9ARB2 linum usita
26	176.5	3.3	1108	10 Q9ARB0	Q9ARB0 linum usita
27	175	3.3	682	4 Q9NVW3	Q9NVW3 homo sapien
28	174.5	3.3	1398	10 Q9C7B4	Q9C7B4 arabidopsis
29	172.5	3.3	1560	10 Q9F7A6	Q9F7A6 arabidopsis
30	172	3.3	968	10 Q9ZTK1	Q9ZTK1 lycopersico
31	171.5	3.2	1110	10 Q9F177	Q9F177 arabidopsis
32	171.5	3.2	1167	10 Q9A471	Q9A471 arabidopsis
33	171.5	3.2	1295	10 Q9FN83	Q9FN83 arabidopsis
34	170.5	3.2	1220	10 Q24015	Q24015 lycopersico
35	170	3.2	1066	5 Q18902	Q18902 caenorhabdi
36	169.5	3.2	1110	10 Q9SDA5	Q9SDA5 arabidopsis
37	165	3.1	1112	10 Q41397	Q41397 lycopersico
38	165	3.1	1112	10 Q41398	Q41398 lycopersico
39	164.5	3.1	1715	10 Q49470	Q49470 arabidopsis
40	164	3.1	1016	10 Q9ZTJ6	Q9ZTJ6 lycopersico
41	163	3.1	512	11 Q9C076	Q9C076 mus musculu
42	163	3.1	657	11 Q9EP67	Q9EP67 mus musculu
43	163	3.1	853	10 Q9ZS82	Q9ZS82 lycopersico
44	162.5	3.1	524	4 Q9BTM6	Q9BTM6 homo sapien
45	162.5	3.1	1119	10 Q9ARB3	Q9ARB3 linum usita

ALIGNMENTS

RESULT 1

ID Q9NPP4 PRELIMINARY: PRT, 439 AA.

AC Q9NPP4:

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)

DE HYPOTHEtical 49.3 KDA PROTEIN (FRAGMENT).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Pluvinet R., Estivill X., Escarceller M., Sumoy L.;

RL Submitted (JUL-2000) to the EMBL/Genbank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA Lehrach H., Poustka A., Lundeberg U.;

RT "The European IMAGE Consortium for Integrated Molecular analysis of human gene transcripts.";

RL Submitted (JUL-2000) to the EMBL/Genbank/DBJ databases.

DR EMBL: AL389934; CAB97523.1; -

KW Hypothetical protein.

FT NON_TER

SO SEQUENCE 439 AA; 49286 MW; 7EFC9C8F0B77359A CRC64;

Query Match 42.5%; Score 2244; DB 4; Length 439;

Best Local Similarity 99.8%; Pred. No. 4.4e-143;

Matches 438; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	586	INSNIPYLFDFEHLNFCASALDFIKLDFYGGAMASWERNAAEDTGGIHMEAPETIYP	645
DB	1	INSNIPYLFDFEHLNFCASALDFIKLDFYGGAMASWERNAAEDTGGIHMEAPETIYP	60
QY	646	SRAVSLFFNWKQOEFTLEVTLRDFSKLNKODITYIGKTFSSATSRLQIKRCAGVAGLS	705
DB	61	SRAVSLFFNWKQOEFTLEVTLRDFSKLNKODIRYIGKTFSSATSRLQIKRCAGVAGLS	120
QY	706	LVLSTCKNIYSIMVSPLETTEDERHITSVNLKTLTIHDLQNGLPGLTDSIGNLKNL	765

```
|||||
Db 121 LVSTKNTISLWVEASPLTIEDERHITSVTNKTLSIHLQNRPLGGFTJSLGNLKNL 180
OY 766 TKLIMONIKMNEDAIKLEAGLNKMKCFHLTHLSIDIGEGMDYIVKSLSSPCDLEI 825
Db 181 TKLIMONIKMNEDAIKLEAGLNKMKCFHLTHLSIDIGEGMDYIVKSLSSPCDLEI 240
OY 826 QLVSCCLSANAAVKILAOINHLNVLKSLDLSENVYLERKDGNEALHELIDRMNVLEOTLALM 885
Db 241 QLVSCCLSANAAVKILAOINHLNVLKSLDLSENVYLERKDGNEALHELIDRMNVLEOTLALM 300
OY 886 LPMGCVQVQSSLSLKLHLEVPOLYVIGLKNMRLTDEIILGAFGKNPLKNPQOLNLA 945
Db 301 LPMGCVQVQSSLSLKLHLEVPOLYVIGLKNMRLTDEIILGAFGKNPLKNPQOLNLA 360
OY 946 GNRVSSDGMVLAIFGVFENLKQVFFDFSTKEFLPDPALVYKLSQVLSKTLFLOEARLVGM 1005
Db 361 GNRVSSDGMVLAIFGVFENLKQVFFDFSTKEFLPDPALVYKLSQVLSKTLFLOEARLVGM 420
OY 1006 QFDDDDLVTITGAFKLVTYA 1024
421 QFDDDDLVTITGAFKLVTYA 439
```

```
RESULT 2
OY1239 PRELIMINARY: PRT: 953 AA.
AC OY239:
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE NOD1 PROTEIN.
GN NOD1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxID=9606;
OX
RN
SEQUENCE FROM N.A.
RA MEDLINE-99262599; PubMed-10329646;
R Medline N., Koseki T., del Peso L., Hu Y., Yee C., Chen S., Carrio R.,
Merino J., Liu D., Ni J., Nunez G.;
RT "Nod1, an Apaf-1-like activator of caspase-9 and nuclear factor-
kappaB."
RL J. Biol. Chem. 274:14560-14567(1999).
RN
SEQUENCE FROM N.A.
RX MEDLINE-99240667; PubMed-10224040;
RA Bertin J., Nir W.-J., Fischer C.M., Tayber O.V., Errada P.R.,
Grant J.R., Kelly J.J., Gosselin M.L., Robison K.E., Wong G.H.W.,
Glucksmann M.A., Distefero P.S.;
RT "Human CARD4 protein is a novel CED-4/Apaf-1 cell death family member
that activates NF-kappaB."
RL J. Biol. Chem. 274:12955-12958(1999).
RN
EMBL: AF149774; AAD43922.1; -
DR EMBL: AF119925; AAD28350.1; -
DR EMBL: AF126484; AAD29125.1; -
DR InterPro: IPR001315; CARD.
DR InterPro: IPR003592; LRR_out.
DR pfam: PF00619; CARD; 1.
DR PROSITE: PS50209; CARD; 1.
DR SMART: SM00114; CARD; 1.
DR SMART: SM00370; LRR; 5.
DR SEQUENCE 953 AA; 107690 MM; 0A9DF5FC6487E21A CRC64;
```

Query Match 5.6%; Score 296; DB 4; Length 953;

Best Local Similarity 20.8%; Pred. No. 2e-11;

Matches 240; Conservative 133; Mismatches 356; Indels 406; Gaps 58;

```
OY 1 MNFKNSRSLALQRMGTIVKQITDLEFVNANVREVNITCEKVEQDARGIHHLK 60
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 20 IOLKSNRELIVHIRNF--QCLVDNLKNDYFSAEDAEIVCACPTPOPKVKRILDILOS 77
```

```
OY 61 KGSSECNLFKSLKENNYPFLFODLQNSLPHQTSBGDLDAQDLK----DIYHPSPFLN 116
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 78 KGESEVSEFVLYLQ-----LADAVYDLRPMLEIGFSPSLT 115
OY 117 FVPLGEDIDIIIFNLKSTFEPV-LMKRQDHH-----RYEQLTL----- 154
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 116 -----QSKVVVN-----DDPVSRYYQOLRHHLRGDSKFEVLYAOKEEILLEIYMDTIM 164
OY 155 -----NGLLQALQS-PCIIIE-----GSGKGRKSTLORIMAMGSGKC 191
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 165 ELVGFSENEISGLNSLACLIDHTTGILNEOGETIFILGAGVKSMLDRDLSMATGRL 224
OY 192 KALTKEFVFEFLRSRA-----QGGLEFTELCDDLDIPGTIRKQTFMAMLTAROR 242
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 225 DAVGVKEFHFRGMFCFCKESDRLCQDILLFKHYC-----YPRDEEVEF-AFLIRFPIV 278
OY 243 VLFLLDGYNEF-----KPNCEPIEALLKENHFRKNMYIVTT--TTECLRIHROG 291
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 279 ALFTFGDLDELHSDLDLSRPVSSCPWEPA-----HPLVILANLISGKLK-----G 325
OY 292 A--LTAEGVDMTEDSAQALIREVLI-----KELAGLLLOQKSRG 330
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 326 ASKLITARTG--IEVPRQPLRKVLLRGFSPSHLRAYARMPERARLORLLSQLEARN 383
OY 331 LRNLKMTPLF--VVITCAIQMGSEFEHSHNQ-----TTFEHPFY-----DLIT 371
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 384 LCLSLCVPLFCWIIIFRC-FQHFRAFEQSPQDPDCTMTFLTVDVLTVEHLNMQSSIV 442
OY 372 QKRNKR-----IKGVAASDFIRSLDHCGLALEGVSHKFPDELQV--SSVNE-DVLLT 423
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 443 QNTRSPVETLHAG--RDTLCSL--GOVAHGRMEKSLVFTEGVQASGLQERMOY- 495
OY 424 TGLCKYTAQRFKPK-----YKFFH-----KSPQEF- 449
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 496 -GFL-----RALPELGPQGDQSYEFPHLTQAFTAFFVLDVAGVQELLRFQEMM 548
OY 450 -TAGRRLSLITSHPEEVTKGNGVYLOKKNVSIISITVYSSLIYTCGSSVEATRAVMKH 508
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 549 PPAGAAATTSCTPFPFLPFCQSGSPAREDFLNKHDFQFTNL--FLCGLLSKAKQKILRH 606
OY 509 LAAVYHGCILGISTAKRPLMRO--ESLOSVKNTTQOELIKAININSF----- 554
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 607 LVPA-----AALRRKRKALMAHLESSLKRGYLS-----LPRQVSEFNOVQOAMPTTMM 655
OY 555 VECGILHLYOESTSKALSQEFQKSLYINSGNIPYLEDFFEHLPNC-----ASA 608
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 656 LRC---IYEQSQKVG--QLAA-----RGICANYL-----KLTYCNACNSADCSA 694
OY 609 LDFIKLDFYGCAMASWEKAEDTGTGIMHEBAPETYIPSR-AVSLFPMKQOEFRTLEVTLR 667
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 695 LSFV-----LH-----HPPKRLADL----- 710
OY 668 DFKLKNKODITVYLGKTFSSATSLRIQKCA-GVAASLSIVSTCKNITISLWVEASPLTI 726
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 711 DNNNLNDYGVRELQPCPSRLTYRLSVNQTDDGCVYLSEELTKYIVTYLGLYNQITD 770
OY 727 EDERHITSVTNKTLSIHLQNRPLGGFTJSLGNLTKLIMDKIKMNEDAIKLAEG 786
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 771 VGARVYTKI-----LDECKGLTHLKGKNTKITSBEGKTLALA 807
OY 787 LKLNKMKCFHLTHLSDIG-----EGMDYIVKSLSSPCDLEIQLVSCCLSANAAVK 838
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 808 VKNSKS-----ISEVGWGNQVGDGAKAFALRNHP-SLTJLTLASNSIGSTEGK 858
OY 839 ILAONLHNLVKSILDLSENVYLERKDGNEALHELIDRMNVLEQTLALMPGCVQVQSSLS 898
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 859 STARALQONTSLIELMLTONLENDVAESLAEML-----KVNO 896
OY 899 LKHLLEVPOLYVIGLKNMRLTDEIIRILGAFGKNPLKN---FOOLAGNVRVSSDGM 955
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 897 TLKHL-----W-LIONQITAKTQALDALALQSNIGTICLNGNIRKEE-- 940
```



```

OY 956 AEMGVFENLKQLVFF 970
Db 941 --AKVYEDEKRITCF 953

RESULT 3
O9H6G6 PRELIMINARY; PRT; 1062 AA.
AC O9H6G6;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CDNA: FLJ22299 FIS, CLONE HRC04653.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxId=9606;
[1]
SEQUENCE FROM N.A.
MA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA Suzuki Y., Ohtsuka M., Nishi T., Shibahara T., Tanaka T.,
RA Nakamura Y., Isigai T., Sugano S.,
RT "NEDO human cDNA sequencing project."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK025952; BAB15293.1;
DR InterPro; IPR000767; Disease_resist.
DR InterPro; IPR003590; LRR_out.
DR InterPro; IPR003592; Disease_resist.
DR PRINTS; PR00364; DISEASERIST.
DR SMART; SM00370; LRR; 7.
DR SMART; SM00368; LRR_RI; 8.
SQ SEQUENCE 1062 AA; 120488 MW; 9F890DF9B7CB944 CRC64;

Query Match 5.0%; Score 266; DB 4; Length 1062;
Best Local Similarity 22.5%; Pred. No. 2.5e-09;
Matches 201; Conservative 123; Mismatches 337; Indels 234; Gaps 42;

OY 166 IIEGSGKSTLLQRIAMLMGSKKALTRKFFVFL--RLSRAOGGLFETLDDLLD 222
Db 210 VLYGAGIGKTTLQAKIMLDAED--NLHKKFAFYLSCRELSRLGCSFAELV--FRD 265
OY 223 IPGTRKOTFAMMLLKLQRYVFLDGYNEFK-----PONCPREI-----ALI 265
Db 266 WP---ELDDIPHLIAQARKILFVIDGFDELGAARGALIEDSCGMKKKPPVVLGSL 322
OY 266 KENHFKMMVYITTECLRIHROFALAE-----VGMTEDSAOALIREVLIKELA 318
Db 333 NKVMLPKAALLVTPRALRLDR---LAEERPIYIRVEGFLEEDRRAYFLRHFGDEDOA 378
OY 319 EGLLQIQKSRCLRLMLTPLFVVITCA---IOMGESEFHSHTOTLTHTFYDLLQKNK 375
Db 379 MRAFLMRSNAALFOLGSAVPCVTCVTLKLOMEKGEDEVPTCLTRGLFRLFCRRP 438
OY 376 HKHKGVAASDFRSLDHGDLALEGVFS---HKFDELQDVSSVNDVLTITGLCKY 430
Db 439 ---GGAQRLGALRTLSL---LAAQGLMAQTVLHREDLERLGVOGSDRLRFLDGIIQOD 492
OY 431 TAQRKPKYKFFHKSFQETVAGRLSLTSHPEPEVTKGNGYLOKMWISIDISTYSSL 490
Db 493 RVS--KGCYSFTHLSFOOF---LTAFLYLTLEKEEEDRQHTW---DIGVQKILSGV 542
OY 491 LR-----YTCGSSVEATRAVMKHLAAYOHGGLLGISTAKRPLMR----- 530
Db 543 ERLRNPDLIQAGYISFGLANE-KRA--KELEATF--GGRMSPTI-KQELRLRDISCKGCH 596
OY 531 -----QESLSQVKNTEQELIKAININSFVCGIHLVQESTSKSALQSEFAFFQGSKL 584
Db 597 STVTDLOELGLGYSQEEBELVKEY-MAQFKELISHLAANDVPPSPCVKHKRNLQKNSL 655
OY 585 YINSQNIID-----YLPDFEHLPCA---SALDPIKL---DPYGG 619
Db 656 QVIRENLEPNVTASESDAEVERSDODDQMLPFWTDL--CSIFGSKKDLMLAINDSFLSA 713

```

```

OY 620 AMAS--WEKAEDTGSIHMEAPETIYIPSRAYSLEFNKQERTLEVTLDROFSKL----- 672
Db 714 SLVRLCEQIASDT--CHLQ-----RVYFNKISPADAHRNLCIALRGHKTIVYTL 762
OY 673 --NKOD-----ITYLGIFFSATSLRLQIKRACVAGSLVLS----- 709
Db 763 QGNDDDDMFALCEVLRHPECLRYLGLVSCSATYQW-----ADLSIALEVNOSL 813
OY 710 TCKNT--YSLMVEASPLTIEDERHITSYTNKTLTSHDLQNRPLGGLTDSIGNIKNLTK 767
Db 814 TCVNLSDELDEGAKLLYTLRHKCF-----LQR 844
OY 768 LIMDNKINMEPAIKLAGLKLKMKCLFHLNL-----SDIGE--GMQYIVKSLSEPPDL 822
Db 845 LSLENCHLLEANKDPLAVALVVSRE---LTHLCLAKNPIGNTGVCFLGRLPYECKL 899
OY 823 BEIQLVSCCLSAVAKIILQNLHNLVKLSIDLSENVLEKDGNEALHBLIDRMVLEQUT 882
Db 900 QTLVLMNCDITSDGCCDLTKLLQEKSSLLCIDLGLNHIGVGMKFLCALARK--PLCNLR 957
OY 883 ALMPFGCDVQG--SLSSLLKHLLEVLPOLVKGLKMMRLTDEIRILGAFPGKNPL 936
Db 958 CLML-WGCSIPPFCSCEDCSALSCNSQSLVTIDL-----GQNPL 994

RESULT 4
O9NX02 PRELIMINARY; PRT; 1062 AA.
AC O9NX02;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE NALP2 (HYPOTHETICAL 120.5 KDA PROTEIN).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxId=9606;
[1]
SEQUENCE FROM N.A.
MA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA Suzuki Y., Ohtsuka M., Nishi T., Shibahara T., Tanaka T.,
RA Nakamura Y., Isigai T., Sugano S.,
RT "NEDO human cDNA sequencing project."
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Martinon F., Hoffmann K., Tschopp J.;
RT "NALP2 a novel NACHT, LRR and PYD containing protein."
RN [3]
RP Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG CARCINOMA;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK000517; BAA91223.1;
DR EMBL; AF310106; AAG30289.1;
DR EMBL; BC003592; AAH03592.1;
DR InterPro; IPR000767; Disease_resist.
DR InterPro; IPR000767; Disease_resist.
DR PRINTS; PR00364; DISEASERIST.
DR SMART; SM00370; LRR; 7.
KW Hypothetical protein.
SQ SEQUENCE 1062 AA; 120514 MW; 4DBB0F6E9C2BC8A7 CRC64;

Query Match 5.0%; Score 264; DB 4; Length 1062;
Best Local Similarity 22.5%; Pred. No. 3.4e-09;
Matches 201; Conservative 123; Mismatches 337; Indels 234; Gaps 42;

OY 166 IIEGSGKSTLLQRIAMLMGSKKALTRKFFVFL--RLSRAOGGLFETLDDLLD 222
Db 210 VLYGAGIGKTTLQAKIMLDAED--NLHKKFAFYLSCRELSRLGCSFAELV--FRD 265

```

OY	223	IPGTRKOTFAAMLLKLRQVLFLLDNEE-----	KPNCP-IEALI	265		
Db	266	WP-----ELDDIPHILAAQARKLLFVIDFGDEGAAPGALIEDICGDWKKKRPVYLSSLL		322		
OY	266	KENRHKFMVIVTTTTECLRHIRQFGALTA-----	VGDMTDSAOALIREVLIKELA	318		
Db	323	NRVMLPKAALLVTTTRPRALRDLR-----	ILAEPIYIRVEGFLIEDBRAYFLRHGDEBQA	378		
OY	319	EGLLLIQIKSKCLRNLMTPLEFVVITCA-----	IQWGESEFHSHTQTLFHFPIYDLLIOKNK	375		
Db	379	MRAPFLMSNAALFQLGSAFAPVCMIVCTTLKIQWKEKEDVPVLTTLTGFLFRLCSRF		438		
OY	376	HKHKVAAADPFIKSDHOGDALGVES-----	HKPFELQDVASVNEVDLITGLGCKY	430		
Db	439	---OGAOLRGALRTLIL-----	LAAGLMAQTSVLRHREDLERAGVESPURLFLOSDILRQD	422		
OY	431	TAQREPKPYKFEFKHSFOEYTAGRRLLSLTSHBEDVYTKGNGYLQKKVMSISDIITYSSL		490		
Db	493	RVS--KCGYSFIHLSQFOF-----	LTFALFYTLKEEEDRQGHFW--	542		
OY	491	LR-----	YTCGSSVATATRAVKHHLAAVYQHCGLLGLSLAKRPLMR	530		
Db	543	ERLRNPDLIQAGYVSFGLANE-KKA--	KELEATP--	GCRRMPDI-KOELLRQDISCKGH	566	
OY	531	-----QESLQSVKNTTPEOELIKANININSFECGITHLYQESTSKALSOEFAFQSKSL		584		
Db	597	STVYDLOGLLCCGLYESQEEELVKEY-MAQFEIHLHLNADVVSFCVAKHCRMLQKMSL		655		
OY	585	YINSGNIPD-----	YLFDFEHLPCA--	SALDFIKL---	DFYGG	619
Db	656	QVIKENLEPENATASESDAEVERSDODDHMLPFMTDL--	CSIFGSKDLMGLAINDSFLSA		713	
OY	620	AMAS--WEKAAEDDGGIHMEAPETIYIPRAVSLFEMKKQEFRTLEVTLADFSL		672		
Db	714	SLVRLCEQIASDT--CHLQ-----	RVEFNKISPADARRMLCALRGHKTVYTL		762	
OY	673	---NKOD-----	ITYLGIKFSSASLRLOIKRGAVAGSLTVLS		709	
Db	763	QGNDDQDFPALCEVLRHPECNLATLGIVGSSATTQOM-----	ADLSLAEVQSL		813	
OY	710	TCKNI--YSLMVEASPLTIEDERHITSYTNLKTLSIHDLQNRLLPGLJDSLGNLKNLTG		767		
Db	814	TCVNLSDNELLDEGAKLLYTLRLRPKF-----			844	
OY	768	LIMDNKNNEDDAIKLGAELNKKMCFHLTHL-----	SDIGE-GMDYIKSLSSPECDL		822	
Db	845	LSELCCHLTENCKDLAALVVSRE-----	LTHCLAKNPNGTGVKFLDEGLYEPCKL		899	
OY	823	EEIQIVASCSLSANAVKILQAINLHLYKLSIDLESNYLEKGNALHLLIDRMNVLEOLT			882	
Db	900	QTVLVMNCDIRISDQCCDITFLQKSSSLICDGLNHIYGVMKFLCEALRK--	PLCNLR		957	
OY	883	ALMPEMGCDVQG-SLSSLLKHLLEVPOLYKLGKLNMRLLTETRIILCAFPCKNPL			936	
Db	958	CLWL-WGCSIFPFCEDLCALSQNSQSLVTLIDL-----	GQNP		994	
RESULT	5					
Q9BVS		PRELIMINARY;	PRT;	1062	AA.	
AC	Q9BVS5;					
DT	01-JUN-2001	(Tremblere, 17, Created)				
DT	01-JUN-2001	(Tremblere, 17, Last sequence update)				
DT	01-JUN-2001	(Tremblere, 17, Last annotation update)				
DE	HYPOTHETICAL	120.6 KDA PROTEIN.				
OS	Homo sapiens (Human).					
OC	Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.					
OX	NCBI_TaxID=9606;					
FN	[1]					
RP	SEQUENCE FROM N.A.					
NC	TISSUE=CHORIOCARCINOMA;					

RA Strausberg R.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases
DR EMBL; BC001039; AH01039.1; -
KW Hypothetical protein.
SQ SEQUENCE 1062 AA; 12057 MW; 4DBD0A7A9C2BC8A7 CRC64;

Query Match	5.0%;	Score 264;	DB 4;	Length 1063;
Best Local Similarity	22.5%;	Pred NO. 3.4e-09;		
Matches 201;	Conservative 123;	Mismatches 337;	Indels 234;	Gaps 42;

OY	166	IIEGSSGKRYLLORIAMLWGSKC	KALT	KRFVEFL	---	RLSRQGGFLFELCOLLD	222				
Db	210	VLYPGAGIKTTLAQKIMDMAED	-	NIHKRYAFLYSLCRELSRLG	PCSPAEVL	--FRD	265				
OY	223	IPGTRKOTFMAMLLKROVUL	FLDGVNEF	-----	KONCPE	-LEALI	265				
Db	266	WE	---	ELODDIPHILIAOARKIL	FVIDGPEBELCAAPGAL	IEDICGDMKKRP	PVYLLGSLT	322			
OY	266	KENHFKPMNVITTTTECGRH	ROGALPAE	-----	VGDMTEBSQAOLIREVLLKELA		318				
Db	323	NRVYMPKAAVLTTTPRALRDL	---	ILAEPIYIRVGFLEEDRRAY	FLRHFGEDDA		378				
OY	319	EGLLIQIOSRCLRNLMKTPLE	VFVITCA	---	IOMGSEPHSHTOYTLFHTFYDLLIOQNK		375				
Db	379	MRAFELMRSMALFOLGSA	PAVCWICTLTKLOMEKGE	DPVPCLTRTGLTRFLC	SRP		438				
OY	376	HHKRYAASDFIRSDHCDL	ALBEVFS	-----	HKDFELODYSSNEVYLLTGILCKY		430				
Db	439	---	QGAOLRGALRTSL	---	ILAAQIMATOSVLAHRDELR	LGVOESDRLFLDGLLRD	492				
OY	431	TAORPKPKYKFFHKSFQOET	TAGRRLSLLTSHSHEPEVT	KGNGVLOKMAV	ISDITSTSYSL		490				
Db	493	RVS	-	KGCYSFHLSEFOE	---	LTALEFYTLKEBEEEDRDHTW	---	DIGVOKLSSV	542		
OY	491	LR	-----	YTCGSSVEATRAVMKHLA	AVOYOGCLLGSIAKRLMR	-----	530				
Db	543	ERLRNPDLIOAGYU	SFGLANE	-KRA	-	KELEATF	-	GCHMSPDI	-	KOELLRCDISCKGH	596
OY	531	-----	QESLOSVKMTTPOGELIKAIN	NINFEVCGIHLVQES	TSKASLSOEFPAFPQGRKL		584				
Db	597	STVTDLOGLGLYSQOEELVKEV	-	MAOFKEISLINA	VDVAVPSEVFCVHCGRNLOKMSL		655				
OY	585	YXNSGNIP	-----	YLFDFEHLPNCA	---	SALDFITKL	---	DFYGG	619		
Db	656	OYIKENLEPENYATASESDA	EVERSODDOHMLP	WPITDL	-	CSIFGSKNDL	MGALINDSFLSA		713		
OY	620	AMAS	---	WEKAADETCGIHMEAPETY	IBSRAVSLFFNMKQEFRTLEV	LTRDSKL	-----	672			
Db	714	SLVRILCEQIASDT	-	CHLO	-----	RYVFKNISPADARHNICL	ALRGHKTVTYLT		762		
OY	673	---	NKDD	-----	ITYLGKIFSSATSRLQIKKACAGV	ASLSVLVS	-----	709			
Db	763	OGNDDDDMFALCEVYL	RHEPCNLRLTGLV	SGCATTOOM	-----	ADLSIALEVNOSL		813			
OY	710	TCKNLT	-	YSIMWEASPLTIEDERHITS	VYNLTSLSHLDQORLP	GGILDSIGNLNTKR		767			
Db	814	TCVNLSDNELDEGAKLL	YTLTRHKKCF	-----						LOR	844
OY	768	LIMDNIKNNEEDA	IKLABGLKMLKMCFL	HTHL	---	SDIGE	-	GMDYTVKSLSEPCD		822	
Db	845	LSLENCHLLEAKCDLA	AVLVYSRE	---	LTHLICAKNPIGNTGKVFCE	GLIRPECKL		899			
OY	823	BEIOLOVSCLSANVAKIT	LAONLHNLVKS	ILDSSENYLEKDNENAL	HELIDRMNVYEOLT		882				
Db	900	QYLVLMWMDITISD	GGCDLTKRLQEKSSSL	CDLJGLNHHGVKMKFCE	SLYLR	-	PLCNLR		957		
OY	883	ALMLPMGCDVOG	-	SLSSLKHLHEEVPOLV	KLGLKMWRLDTFIRILGAF	FGKNPL		936			
Db	958	CIML	-	WGCSIPFSCEDLCSAL	SCNOSLVTIDL	-----	GNPL		994		

Q9HAV9 PRELIMINARY; PRT: 1033 AA.
 ID Q9HAV9
 AC Q9HAV9
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE NUCLEOTIDE-BINDING SITE PROTEIN 1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NX NCBI_TaxID=9606;
 RN [1]
 RA SEQUENCE FROM N.A.
 RA Bertin J., Distefano P.S.;
 RT "The PRIN Domain: A Novel Motif found in Apoptosis and Inflammation
 RT Proteins.";
 RT Cell Death Differ. 0:0-0(2000).
 DR EMBL; AF298547; AAG15253.1;
 DR InterPro: IPR000767; Disease_resist.
 DR InterPro: IPR003592; LRR_out.
 DR InterPro: IPR003590; LRR_RNinh.
 DR PRINTS; PR00364; DISEASERLIST.
 DR SMART; SM00370; LRR_6.
 DR SMART; SM00368; LRR_R1; 8.
 SQ SEQUENCE 1033 AA; 117387 MW; C16152AD510BCA63 CRC64;

Query Match 5.0%; Score 261.5; DB 4; Length 1033;
 Best Local Similarity 22.5%; Pred. NO. 4.8e-09;
 Matches 201; Conservative 123; Mismatches 336; Indels 235; Gaps 43;

166 IIEGSGKSTLLQRIAMWGSGKALTKPKFVFL---RLSRAOGGLPETLDDLLD 222
 Db 182 VLYGAGAGIKTTIAOKMLMDNED--NIHKFYAFYLSCHRELSLGGCSRAELY--FRD 237
 Qy 223 IPGTRKOTFAMMLIKLQRYVFLLDGYNEF-----KQNCPE-IEALI 265
 Db 238 WP---ELDDIPHILAAQRIKILFVDFGDELGAARGALIEDICGDMKKRPVYLLGSL 294
 Qy 266 KENHFKMVIYTTTTECLRNHROGALTAE-----VGMTEPSAALIREVIRELA 318
 Db 295 NKVMLPKAALLVTTTPRALRDLR---ILAEPIYIRVEGLEEDKRAVYLNHFGDEDOA 350
 Qy 319 ECLLIQIOKSRCLRYMLKTPLEVVITCA---IOMGESEFHSHTQTLHTTFYDLILQKNK 375
 Db 351 MRAFLMNSNALRFDGSAAPAVCAIVCTTKLQMEKGEDPVPTCTTRGCLFRLPCLSRFP 410
 Qy 376 HKHKGVAASDFIRSLDHGDLAEGVFS---HKFDELODVSSVNDVLLTGGILCKY 430
 Db 411 ---QGAQIRGALRTLSTL---LAAQGLMAQTSVLRHREDLERLGVQESDLRLFLDGLIRQD 464
 Qy 431 TAQRKPKYKFFHKSFQETYACRLSLTSHPEEVTKGNGYLOKMSISDITSTYSSL 490
 Db 465 RVS--KGYSYFHLSPFOF---LTALEPYLKEKEEDRDGHTV---DIGVQKLSLV 514
 Qy 491 LR-----YCGSSVEATRAVMKHLAAVYOHGCLGLSIAKRPMLR----- 530
 Db 515 EKLRLPDLIQAGYISFGLANE--KRA--KELEATF--GCRMSDIT--KQELIRCDISCKG 568
 Qy 531 -----QESLOSVKNTTQELIKAININSFVCGIHLVQESTSKALSQEFAPFGKSL 584
 Db 569 STVLOLQELGLGYESQEBELVKEV--MAQFKRISLHLNADVVPSPFCVKKHCRNLQKMSL 627
 Qy 585 YINSQNTD-----YLPDFEHLPCA---SALDPFKL---DFYGG 619
 Db 628 OYIKENLEPNVTASESDAEVERSDODQHLPFWTL--CSIFGSKKIDMGLAINDSFLSA 685
 Qy 620 AMAA--WEKAADTGCIHMEAPETIYIPRAVSLFNNMKQEPRTLEVTLRDSKSL----- 672
 Db 666 SLVRLICEQIASDT--CHIQ-----RYVFKNISPADAHRNCLALRGKITYTYITL 734
 Qy 673 --NKOD-----ITYLGISSATSRLQIKRCAGVAGSLSLVLS----- 709

Db 735 QGNDQDMFPALCEVLRHPECNLYRLGLVSCSATTOGW-----ADSLALEVQSL 785
 Qy 710 TCKNI--YSLWEASPLTIEDERHITSVTNKTISTIHDLQNRGLTSLGNLTK 767
 Db 786 TCVNISDELDEGAKLLYTLRHKCP-----LQR 816
 Qy 768 LIMNINKNEEDAIKLAEGLNKKMCLFHLTHL-----SDIGE-GMDYVKSLSSEPCDL 822
 Db 817 LSLKCHLTENCMCDIAVLVRSR-----LTHICLAKNPIGNNGVFLCGRLYPECKL 871
 Qy 823 ERIOLVSCSLANAVKILQNLHNLVKISILDSENYLEKGDNALHEDIRNAVLEOLT 882
 Db 872 QTLVLMNCIDTSDCCDITLKLOEKSSILCIDLGLNHIGVGMKFLCEALRK--PLCNLR 929
 Qy 883 ALMLPWGCDVYG--SLSSLKHLLEVPOLYKGLGNWRITDIEIILCAFPCKNPL 936
 Db 930 CLWL-WGCSIPFSCEDICSLSN-QSLVTLIDL-----GQNP 965

RESULT 7

Q9R1M5 PRELIMINARY; PRT: 1111 AA.
 ID Q9R1M5
 AC Q9R1M5
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE OOP/LAM SPECIFIC PROTEIN.
 GN MATER OR OPT.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 NX NCBI_TaxID=10090;
 RN [1]
 RA SEQUENCE FROM N.A.
 RA STRAIN=NIH/SWISS;
 RA Tong Z.-B., Nelson L.M.;
 RT "A mouse gene encoding an oocyte antigen associated with autoimmune
 RT premature ovarian failure.";
 RL Endocrinology 0:0-0(1999).
 DR EMBL; AF074018; AAB51762.1; -.
 DR HSSP; P10775; 2BNH.
 DR MGD; MGI:1345193; Mater.
 DR InterPro: IPR000767; Disease_resist.
 DR InterPro: IPR001611; LRR.
 DR InterPro: IPR003592; LRR_out.
 DR Pfam; PF00560; LRR_2.
 DR PRINTS; PR00364; DISEASERLIST.
 DR SMART; SM00370; LRR_7.
 SQ SEQUENCE 1111 AA; 125501 MW; 4016A5D67A1C01F4 CRC64;

Query Match 4.6%; Score 242; DB 11; Length 1111;
 Best Local Similarity 20.0%; Pred. No. 1.1e-07;
 Matches 210; Conservative 151; Mismatches 395; Indels 296; Gaps 46;

143 DOHHIRVQLTLNGLQLOS-----PCIIEGSGKSTLQRIAMWGSGKCKALTKF 197
 Db 166 DLHYSPKMLSLDAFKYQKTFQDPHTIILGKRGVGSALARSIVLGMAG--KLFQKM 223
 Qy 198 KVFVEL-----RLSRAOGGLPETLDDLLDIPGTRKOTFAMMLIKLQRYVFLLD 248
 Db 224 SFVIFSVREIKWTKSSLAQ--LIAKRCPSWDLVTKIMSQP-----ERLLFVID 272
 Qy 249 GYN-----EFKPNCPET--EALIKENHFRNMVIYTTTTECLRHIOFG 291
 Db 273 GLDDMDSVLQHDMDTLSDHMDQEPYITLWYSLRKALLPSPFLITTRNRTGLEKSM- 331
 Qy 292 ALTAEVGMDTDSQAQLREVILKELAGLLQIOKSRCLNLT----- 334
 Db 332 -----VSPLYI--LVGCLASRSQSLVLEINISNESRIOVPHSLIENH 373
 Qy 335 -----MKTPLFVVTIC--AIOMGES-----EFHSHQTQTLFHT--FYDLLIQKNKHKG 381

Db 374 QLFDCQAPSVCSLVEBALQLOKKLGRCTLPQQTLTGLVATLVFHOULTK--RPSQAL 431
 382 AASDFIRSLDHCGLALEGVSHK---FDELDQVSSVNEVDLLTTGL-LCKYTAQRFKP 437
 Db 432 SOEEOQTIVGLC--MMAEGVWTRSVFYDDDLKNYSIKESIELAFHNNILLQVGHNSQ 430
 438 KVFHFHSFOEYTAGRR--LSSLLTSHPEEYTKNGYLOKMSISDITSTYSSL---- 491
 Db 491 CVYFHSLSLQDFFAALYYVLEGL-----EEMNHCFEINQRSIMEVKRTDTRLGKMK 544
 492 RYVCGSSVETRAVMKHLAAVYOHGCLGLSTAKRPLMRQESLOSQKNTQELKAINI 551
 545 RLFLFGL---MKNKDLTKLEVLFEYPIVPTVE-OKLOHWVSLAQOVNGTSPMDTLDF-- 558
 552 NSFEEGCHLYOESTSKSALSOEFAFQGSXYINSQ---NIPDYLFDFEHLFPCAS 608
 599 ----YCLFESQDEEFGALKR-----FQEWLLINQKMDLKVSSC-----LKHQON- 642
 609 LDTIKLDFYGGAMASWEKAADTGGIHMEAPETYPISRAVSLFPNNKQEFRTLEVTLD 668
 643 LKAIRVDIRD--LLSYDNTLELCPVYTVQV-----TQCKPLMEWMGNFCVLSLRN 693
 669 FSKLNQD-----ITVGLKTFSSATSLR-L 692
 694 LKELDGLDLSLSORAKKILCLELRNOSCRQKLTFSAEVYSGIKHLKLLFSNQLKYL 753
 693 QIKRCAGVAGSLSVLSTCK---NIYSLMVEASPLTIEDERHIT---SVTNLKTLSIH 744
 754 NLGNTPMKDDMKLACEALKHPKCSVETLRDLSCGLTITIGEMISTLLSTTRKLS-- 811
 745 DLONQRLPGSLTSLGNL-----KNLTKLIMD 771
 812 -LAKNRVGSMSLSTGALSSSMCLQKLLILDCGLTPASCHLLVSALFSONMLTHLCLS 870
 772 NIMNEEDATKLAGLKN---LKKMCLFH-----LTHLS-----D 803
 871 NNSLGTBQVQLOFLNPPCALQRLLNHCNIVDAYSGLAMRLANNTRKLTLSLTMP 930
 804 IGEG-MDYIKSLSEPCDEEIQVSCSLANAVKILAEONLNLVSLTLDSENYLEK 862
 931 VGDGAMKLCEALKEPCTYQLELELVQCTQNCCEDLACMITTKTKLKSLDGNNAIGD 990
 863 DQNEALHELDRKNNLEQLTALMLPMGCDVQGSLSLKLLEVPOLVKLGKNNRLTDT 922
 991 KGVITL-----CEGLKSSSSLR-----RLGLACKKLTSN 1020
 923 EIRILAFKPNLKNFOQLNLAGNRVSDGMLAFMGVFE-NIKOLVFPDSTKEFLDP 981
 1021 CCRALSLAISCNF--HNSLNLVKNDFSTSGMLKLSAFQCPVSNLGIITGKQERY--- 1075
 982 ALVRLKLSOVLKSLTLQEARLV--GMQFDDDD 1011
 1076 ARVRR---QLEEVYFVKPHVYIDGDWYASDED 1104
 RESULT 8
 09JLR2 PRELIMINARY; PRT; 1111 AA.
 AC 09JLR2;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE MATERNAL-ANTIGEN-THAT-EMBRYOS-REQUIRE PROTEIN.
 GN MATER.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxId=10990;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129/SV; TISSUE=OOCYTE;
 RX MEDLINE=20222768; PubMed=10754103;
 RA Tong Z.B., Nelson L.M., Dean J.;

RT "Water encodes a maternal protein in mice with a leucine-rich repeat
 RT domain homologous to porcine ribonuclease inhibitor."
 RL Mamm. Genome 11:281-287(2000).
 DR EMBL; AF143573; AAF64393.1; -.
 DR EMBL; AF143559; AAF64393.1; JOINED.
 DR EMBL; AF143560; AAF64393.1; JOINED.
 DR EMBL; AF143561; AAF64393.1; JOINED.
 DR EMBL; AF143562; AAF64393.1; JOINED.
 DR EMBL; AF143563; AAF64393.1; JOINED.
 DR EMBL; AF143564; AAF64393.1; JOINED.
 DR EMBL; AF143565; AAF64393.1; JOINED.
 DR EMBL; AF143566; AAF64393.1; JOINED.
 DR EMBL; AF143567; AAF64393.1; JOINED.
 DR EMBL; AF143568; AAF64393.1; JOINED.
 DR EMBL; AF143569; AAF64393.1; JOINED.
 DR EMBL; AF143570; AAF64393.1; JOINED.
 DR EMBL; AF143571; AAF64393.1; JOINED.
 DR EMBL; AF143572; AAF64393.1; JOINED.
 DR MGD; MGI:1345193; Mater.
 DR InterPro; IPR000767; Disease_resist.
 DR InterPro; IPR001611; LRR.
 DR InterPro; IPR003592; LRR_out.
 DR Pfam; PF00560; LRR; 2.
 DR PRINTS; PR00364; DISFASERST.
 DR SMART; SM00370; LRR; 7.
 DR SEQUENCE 1111 AA; 125544 MW; B07F9DF1B18DD855 CRC64;

Query Match 4.6%; Score 241; DB 11; Length 1111;
 Best local Similarity 19.5%; Pred. No. 1,3e-07;
 Matches 204; Conservative 158; Mismatches 399; Indels 286; Gaps 44;

143 DQNHREQLTGLGLQALQS-----PCIEGESGKSTLLQRIAMLSGSKALTRF 197
 166 DLHYDPEPKMLSDAKRPQKTFQPHITLHGRPGVSKALASIVLGMAQG--KLFGKM 223
 198 KEVFFRLRSLR---AAGLFETLCDQLDIPGTRKQTFAMLLKRLQRYLFLDGYN-- 251
 224 SFVIFFSVREIKWTKESSLAQLIAKCPDSDPVTK-----IMSOPERLFYIDLDIM 277
 252 -----EPRKPNCPRI--EALIKENHRFKNVYVTTTCTELRHIRQFALTA 296
 278 DSVLQHDMDTLSDMKDDEQPTIYILMSLRKALLPDSFLITTRTGLEKLSM----- 331
 297 VGDMDTDSQALIREVLIELAEGLLQIQSKRLNL----- 334
 332 -----VVSPLYL--LVGGLASRRSQVLLENISNESDRIGVFSLIENHQLFDQ 378
 335 MKTPFLVVITC-AIQNGES-----EFHSHTQTLFHT--FYDLLQKNNKHKGVAA 386
 379 QCAPSVCSLVEBALQLOKKLGRCTLPQQTLTGLVATLVFHOULTK--RPSQALSQEQ 436
 387 IRSLDHCGLALEGVSHK---FDELDQVSSVNEVDLLTTGL-LCKYTAQRFKPKFF 442
 437 ITLVGLC--MMAEGVWTRSVFYDDDLKNYSIKESIELAFHNNILLQVGHNSQCYVS 495
 443 HKSFOEYTAGRR--LSSLLTSHPEEYTKNGYLOKMSISDITSTYSSL-----RYTG 496
 496 HSLDQEFALYYVLEGL-----EEMNHCFEINQRSIMEVKRRTDTRLGKMKFFLG 549
 497 SSVEATRAVMKHLAAVYOHGCLGLSTAKRPLMRQESLOSQKNTQELKAININSFVE 556
 550 L--MKNKDLTKLEVLFEYPIVPTVE-OKLOHWVSLAQOVNGTSPMDTLDF-----Y 599
 557 CGHILYQESTSKSALSOEFAFQGSXYINSQ---NIPDYLFDFEHLFPCASALDFIK 613
 600 CLFESQDEEFGALKR-----FQEWLLINQKMDLKVSSC-----LKHQON-LKAIR 647
 614 LDFYGGAMASWEKAADTGGIHMEAPETYPISRAVSLFPNNKQEFRTLEVTLRDSKIN 673
 648 VLIRD--LLSYDNTLELCPVYTVQV-----TQCKPLMEWMGNFCVLSLRNLKELD 698
 674 KQD-----ITVGLKTFSSATSLR-LQIKRC 697

```

Db 699 LGDSTLSORAMKILCELEHNSCRIOKLPKSAEYVSGIKHLMKLFSSNQNLKYLNGMT 758
Oy 698 AGVAGSLSTVLSTCK-----NIVSIMWEASPLTIEDERHIT-----SYTNKLTISIDLOQ 749
Db 759 PKKDDMKLACALHPKCSVETLRIDSCETLTIGYEMISTILLISTRLKLS---LAKN 815
Oy 750 RLPGGLTDSLGNL-----KNLTKLIMDNKKN 776
Db 816 RGVGVSMTISLGNALSSMKCLLKLINDNGLTPASCHLLVSALFSSQNTLTHCLSNISG 875
Oy 777 EEDAIKLAGLKN---LKKMCLFH-----LTHLS-----DIGEG- 807
Db 876 TGVGOOLCOFLRNPCALQRLILNHCNIYDVGFLAMRLANNKTLHLSTLMNPYGDA 935
Oy 808 MVIYKSLSSPEODEEIDOLVSCCSANAVKTLAONLHLVLSIIDLSENLEKDNNA 867
Oy 936 MKLCEALKEPTCYLOELVDCQTLQNCEDLACMTTITTKQLKSLDGNMLGKGVIT 995
Oy 868 LHELIDRMNVLEQLTALMLPMGCDVOGSLSLKHLLEVPOLVKLGKRWRLDTEIRIL 927
Db 996 LCEGLKQNN-----SSLRRLGCGACKLTSNCCAL 1025
Oy 928 GAFEGKNPLKNFOCLNAGNRVSSDGWLAFMGVFE-NLKQVFFDFSTKEPLPDALYRK 986
Db 1026 SLAISQNP--HNSINLVKNDSTSGMLKCSAFQCPVNLGITGLMKQEY---ARVRR 1080
Oy 987 LSOVLSKLTFLQDARLY--GQFDDDD 1011
Db 1081 ---OLEEVEFKPHVYIDGMYASDED 1104

RESULT 9
O9HC29 PRELIMINARY: PRT: 1040 AA.
AC O9HC29;
DT 01-MAR-2001 (TREMBLREL. 16, Created)
DT 01-MAR-2001 (TREMBLREL. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLREL. 17, Last annotation update)
DE NOD2 PROTEIN.
CN NOD2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
Ogura Y., Inohara N., Benito A., Chen F.F., Yamaoka S., Nunez G.;
"Nod2, a Nod1/Apa1 family member that is restricted to monocytes and
activates NF-kappaB."
J. Biol. Chem. 0:0-0(2001).
EMBL: AF178930: AAC33677.1;-.
InterPro: IPR001687; ATP-GTP_A.
InterPro: IPR001315; CARD.
InterPro: IPR001611; LRR.
InterPro: IPR003592; LRR_out.
InterPro: IPR003590; LRR_RnInh.
Pfam: PF00560; LRR_2.
DR SMART: SM00114; CARD.1.
DR SMART: SM00370; LRR.5.
DR SMART: SM00368; LRR_RI.7.
DR PROSITE: PS50209; CARD.1.
SQ SEQUENCE 1040 AA; 115282 MW; 0037592D96D7D0FF CRC64;

```

Query Match 4.4%; Score 234.5; DB 4; Length 1040;
 Best Local Similarity 20.1%; Pred. No. 3.2e-07;
 Matches 214; Conservative 157; Mismatches 384; Indels 311; Gaps 42;

Oy 4 IKDNRALIOBGMVYVIOITDLDLVNVLNREVNIIICCEVE-----QDAAGCIHM 57
 Db 134 LOSHRPAIVRLHSHVENKL--DL-AW--ERGFVSQYECDEIRLPIPTPSQARRRLDL 187

```

Oy 58 ILKGSSECNLFKSLKEMNYPFLQDL-----NGOSLFHOTSEGD----- 97
Db 188 ATVKANGLAFLFLQIVQELPVLALPLEAATCKKTMALRTVSAQSFSTYDAEFLC 247
Oy 98 LDDLAD-----LKDYHTPSFLNFPYLGSDIDIIIFNLKSTF 134
Db 248 LEDIYENLVEMADVGMAGPQKSPATLGLDELESTFGHLN-----DDADTV----- 295
Oy 135 TEPVLMRKDQHHRVLEQLTNGLLDALOSPCLIEESGKSGSTLLORLAMLGSGKCAL 194
Db 296 -----LVGEASGKSTLLQRLHLMAAG--ODF 322
Oy 195 TKKFFVFLRLSRAO-----GGLFETLQDOLDIPGTRKOTFMAMLLKLRLQVLF 245
Db 323 QEFLVFPFSCQLOCMKPLSVRTLLEHCCW-----DVGQDITQLLDHPRVLL 376
Oy 246 LLDGYNEFK-----PONCEIEAL--IKENHRFN--VYVITTECLRHIRQ 289
Db 377 TDEGDFEEFRTDRERHCSPTDPTSVQTLFNLQNLKARKRVNLSRPAVSAFLRK 436
Oy 290 FGALTAEGVDMEDSAQALIREVLKE--LAEGLLIQIOKSRCLRLDMTPLE--VITICA 346
Db 437 YIRTEFNLKGFSEOGIEYLRRHHEPGVADRLIRLOETSLMGLCHLPVFSMWVSKCH 496
Oy 347 IOMGESEFHSHTQTLFTEFYDLLIOKNKH-----KGAAS--DFIRSDHCGD 395
Db 497 QELLQEGGSPKTTTDM-----LLIOHFLHATPPDSASQGLSGLRGLRPTLIHLGR 552
Oy 396 LALEG-----VSHKFEFELQDVSSVNEVDLLTGLLCKYTAQRFKRYKFFHKSFOEY 449
Db 553 LALMGLGMCYVFSQAQ--QLQAAQVSPDDISLGLFVRAKGVPGSTAPLEFLHTTFQCF 609
Oy 450 TAGRLSLSLTSHPEEYTK-----GNGLOKMNYSID-----TSTYSSILRTIC 495
Db 610 FAAYFLA--LSADVPPALLRHLENCGRPNSPMALRLPTMCIOASEGKSSVAALQAAE 667
Oy 496 GSSVETRAVWKHLAAVYOHGCLLGLSLAKRPLRQESLQSVKNTTEQEIILKAININSFV 555
Db 668 PHNLQITTAFLAGLSRHEWGLLACQTSERALLRQA----- 705
Oy 556 ECGIHLVQESTSKSALSGEFAEFQGKSLYINSQNIPIYLFDFEHLPCASALDFIKLD 615
Db 706 -CARMLARSLKHHHSIPPAARGEAKSVHAMPG-----FIWLI 743
Oy 616 FYGAMASMEKAEDTGGIHMEARETYI--PSAAVSLFFEMKQEPRTLEVTLRDFSKL 672
Db 744 RSLVMEERLARKAARGLVNGHLKLTFCVSGPTCCALAFVLQHLRRVALQD----- 797
Oy 673 NKODITTYLGIKIFSSATSLRLQIKRCAGVAGSLVSTCKNIY--SLAYEASPLTIED 728
Db 798 ---DYSVGDICVE-----QLLPCLGV-----CKALYLRDNNISDRGICKLIEC 838
Oy 729 EKHITSVTNLKTLSIHDLONLPGGLTDSLGNL-----KNLTKLIMDNKKNEDAIKLA 784
Db 839 ALHCEQLQKLA-----LFNNKLTGCAHSMAKILACQNFALRLGNNYITAAQAYLA 892
Oy 795 EGIKMLKKMKLF-----HLTHLSDIGGGMVYIKYKSLSSPECD 821
Db 893 ECLRGNTSLDQFLGFWGNRVGDEGAQALAEALDQSLMELSVGN--TGSVGAQOLA 948
Oy 822 L---EELIVSCSLSAN-----AVKILAONLHNLVLSIIDLSEVYLEKDEGNALHEHID 873
Db 949 LMLAKNVMLEELCELEENILQDEGVCSLADGLKKNSSKLTILKSNNICITYLGAENALQALE 1008
Oy 874 RKNVLEQLTALMLPWGCDVOGSLSLKHLLEVPOLVKLGKKNWL 919
Db 1009 RNDTILEV-----W---LRGNTFS---LEEVD---KLGRDRTFL 1038

```

RESULT 10
 O9HAV8 PRELIMINARY: PRT: 1429 AA.
 AC O9HAV8;

DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, last annotation update)
 DE CASPASE RECRUITMENT DOMAIN PROTEIN 7.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID=9606;
 RN
 RN SEQUENCE FROM N.A.
 RA BERTIN J., DISTEFANO P.S.;
 RT "The PYRIN Domain: A Novel Motif found in Apoptosis and Inflammation
 RT Proteins.";
 RL Cell Death Differ. 0:0-0(2000).
 RP
 RP SEQUENCE FROM N.A.
 RA BERTIN J.;
 RL Submitted (AUG--2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF298548; AAG15254.1; -;
 DR InterPro: IPR000767; Disease_resist.
 DR InterPro: IPR001611; LRR.
 DR InterPro: IPR003592; LRR_out.
 DR InterPro: IPR003590; LRR_Rninh.
 Pfam: PF00560; LRR_1.
 PRINTS: PR00364; DISEASESISIT.
 SMART: SM00370; LRR; 5.
 SMART: SM00368; LRR_R1; 5.
 SQ SEQUENCE 1429 AA; 160880 MW; 90F809AB838B7F7 CRC64;

Query Match 4.18; Score 217; DB 4; Length 1429;
 Best Local Similarity 20.9%; Pred. No. 7.8e-06;
 Matches 185; Conservative 97; Mismatches 289; Indels 314; Gaps 36;

QY 142 KDQHHNRVEQLTNGLLQALOSP--CIIEESGSGKSTLQRIAMLGSGK----- 190
 DB 305 EENRGHILRIQLFGLDQPEPRIVILQGAIGKSTLARKQKAEAGRGQLDGRFQHV 364
 QY 191 -----CAALTKEFVFLLRLSRAOGLFETLCDLLDIPGIRKOTFAMLLKRLQRYLFL 246
 DB 365 FYFSCKRLAQSKVYSLAELGKDG-----TATPAPIRQ-----ILSRPERLLFI 408
 QY 247 LDGYNKRPONCEIELEIKENHFRKMMVITTTTECLRIHROFALTAEGVDMTEDSAQ 306
 DB 409 LDGVDL-----PGWVLOEPPSSSELCLHWSQ-----POPAD 437
 QY 307 ALIREVLIKE-LAEGLLLOQSKRLNKKTPLEVVITCAIQMGSEFHSHTOTLEFHT 365
 DB 438 ALLGSLGKTIPLPASPILITARTALONLI--PSLEQARWVEVLGSE--SSKREYFRY 493
 QY 366 FYD-----LLIQNKHKHKGVAADFIRSL-- 390
 DB 494 FTDERQAIRAFRLVYSNKKELMALCLVPVWSMLACTCLMOQKRRKELTLNLSKTTTCLH 553
 QY 391 -----DHGDDALEGVFSHKPFDELQDVSSVND-----VLTTGLLCK 429
 DB 554 YLAQALQALQPLQRLQPLC-SLAEBGIMORKTLLFSPDLKRGDLGAIISTFLKMGILQ 612
 QY 430 YTAQRFKPKKFFPKSKPOETAGRRLLSLTSHPEEVTGNGVQLQMWISDITSTYSS 489
 DB 613 HP--PLLSFTHLCQDEFFAA--MSYVL-----FDEKGRG----- 644
 QY 490 LLRYTGSSEVATRAVKNHLLAAYVQHGLGLSLAK--RPLMROESLQSVKNTTEQELK 547
 DB 645 -KHSNCLIDEKT-----LEAVGIGH-LFGASTFRFLGLLSDGREGMENIFHCRLS 696
 QY 548 AININSVEEC-----GIHLVQESTKSALSOEFAPFOGKLYINSGN---IPD 593
 DB 697 GRNIMQWVPSIQLDLQPHSLESLSLCLYETRNKTFLLQVM-AHFEEMQMCEVTDMLLCT 755
 QY 594 YLDFEFELHNCASALDFIKLDFYGGAMASMEKAAEDTGGIHMEAEATVIPSAAVSLF- 652
 DB 756 FCIKFSNHRV-----KLQLEIG-----RQHRSTSPSKMV-LFR 788

QY 653 -----FNNKOEFTLEVLRLDESKLNKODITYLGIKIFSSATSLRLQIKKACVAGSL 706
 DB 789 WVPYDAVWQILFSLVLT-KNLKELD-----LSQNSLSHSA 824
 QY 707 VLSTCKNITSVMWASPLTIEDERHITSVNLKTLSTIHDLOQRPLG-GITDSLGNIKNL 765
 DB 825 VKSLCK-----TLRRPRCL-----LETRLAGCGILT 850
 QY 766 TKLIMONIKMNEBDAIKLAGLKNKKMCLFHLJHLSDIGEGMDYIKSLSPDCLEET 825
 DB 851 -----AECKDLATGLRANOTLTLELDSFNVLMAGAKHLQRLRPSCKLQRL 899
 QY 826 QLVSCCLSANAVKILAQNHLVKLSTLDSFNVLEKDGNEALHELIDRANVLEQLTALM 885
 DB 900 QLVSCGLTSDCCODLASVLSASPSLKELDLQGNNDVGVRLCE----- 944
 QY 886 LPMGCDVQSGSLSLKHLEVPOLVGLKKNRLTD---TEIRIL 927
 DB 945 -----GLRH--PACRLIRGLDQTLTLDSEMRQELRAL 974

RESULT 11

ID Q9UFT4 PRELIMINARY; PRT; 1192 AA.
 AC Q9UFT4;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, last annotation update)
 DE HYPOTHETICAL 134.6 KDA PROTEIN (FRAGMENT).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID=9606;
 RN
 RN SEQUENCE FROM N.A.
 RA TISSUE-OTERUS;
 RC Koehler K., Beyer A., Mewes H.W., Gassenhuber J., Wiemann S.;
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL117470; CAB55945.1; -;
 DR InterPro: IPR000767; Disease_resist.
 DR InterPro: IPR001611; LRR.
 DR InterPro: IPR003592; LRR_out.
 DR Pfam: PF00560; LRR; 2.
 DR PRINTS: PR00364; DISEASESISIT.
 DR SMART: SM00370; LRR; 5.
 KW Hypothetical protein.
 FT NON_TER 1
 SQ SEQUENCE 1192 AA; 134610 MW; D28D00457C36229A CRC64;

Query Match 4.0%; Score 212; DB 4; Length 1192;
 Best Local Similarity 20.8%; Pred. No. 1.3e-05;
 Matches 184; Conservative 97; Mismatches 290; Indels 314; Gaps 36;

QY 142 KDQHHNRVEQLTNGLLQALOSP--CIIEESGSGKSTLQRIAMLGSGK----- 190
 DB 24 EENRGHILRIQLFGLDQPEPRIVILQGAIGKSTLARKQKAEAGRGQLDGRFQHV 83
 QY 191 -----CAALTKEFVFLLRLSRAOGLFETLCDLLDIPGIRKOTFAMLLKRLQRYLFL 246
 DB 84 FYFSCKRLAQSKVYSLAELGKDG-----TATPAPIRQ-----ILSRPERLLFI 127
 QY 247 LDGYNKRPONCEIELEIKENHFRKMMVITTTTECLRIHROFALTAEGVDMTEDSAQ 306
 DB 128 LDGVDL-----PGWVLOEPPSSSELCLHWSQ-----POPAD 156
 QY 307 ALIREVLIKE-LAEGLLLOQSKRLNKKTPLEVVITCAIQMGSEFHSHTOTLEFHT 365
 DB 157 ALLGSLGKTIPLPASPILITARTALONLI--PSLEQARWVEVLGSE--SSKREYFRY 212
 QY 366 FYD-----LLIQNKHKHKGVAADFIRSL-- 390

```

DB 213 FTDEROAIARFLVKSNEKELMALCLVPWVSWLACTCLMOQMKREKLTSTTTTLCIH 272
OY 391 -----DHGDLALGVESHKRFDELQVSSVNE-----VLTITGLICK 429
DB 273 YLAQALQAPLGPOLRDLCL-SLAAGIWKKTLPSPDDLKRGHGLGAIITFLKMGIIQE 331
OY 430 YTAQRFKPKYKFFKSFQETAGRLSLTSHPEEYTKNGYLQKWSISDITSTYSS 489
DB 332 HP---IPLSYFIHLCFOEPFAA--MSYVL-----EDEKGRG----- 363
OY 490 LIRYCGSSVEATRAVMKHLAAVYOHGCLLSIAK--RPLMRQESLOSVKNTPEOELIK 547
DB 364 -KHSNCIIDELEKT-----LEAYGIHG--LFGASTTRFLGLISDGEREMENIFHCRLSQ 415
OY 548 AININSFEVC-----GIHLQESTKSALSOEFAFQOKSLYINSGN---IPD 593
DB 416 GRNLQWQWPSIQLLQPHSLHCLYETRNKTFLOVA--AHFEEMGCVETDMELVCT 474
OY 594 YLDFEFELPNCASALDFIKLDFYGGAMASWEKAEDTGIMBEAPETYPISRAVSIF- 652
DB 475 FCIKFSRHVK-----KIQLEIG-----RQHRSTWSPTMVY--LFR 507
OY 653 -----FMWKEFRTLEVTLRDFSKLNODITYLGIKIFSSATSLRLQIKRCAGVAGSLSL 706
DB 508 WVPYTDAYWQILFSLVAKYT--RLKKELD-----LSGNSLSHSA 543
OY 707 VLSTCKNIYSILMVEASPLTIEDERHITSVTNLKLTSHDLQONRLPG--GLTDSLGKLNKL 765
DB 544 VKSLCK-----TLRRPCL-----LETLRLAGGGL----- 569
OY 766 TKLIMDNIKMNEDAIKLAEGKLNKMKCLFHLTHLSDIGEMDVIKSLSEPCDLEI 825
DB 570 -----AEDCKDLAFGLRANQTLTDLSPNVLTDAKAKHLCORLPSPCKLQRL 618
OY 826 QLVSCCLSANAVKLLAONLHNLVKLSILDSENYLEKDGNEALHEDRMVNLQDLTALM 885
DB 619 QLVSCGLTSDCCODLASVLSASPSLKELDLQONLDDVGVRLCE----- 663
OY 886 LPMGCDVQSGSLSLKHLKEEYQVLKLGKMKRLTD---TERIL 927
DB 664 -----GLRH--PACKLIRLGLDQTLTSDMKRQELRAL 693

```

RESULT 12

OY12E0 PRELIMINARY: PRT: 1429 AA.

OY12E0:

01-NOV-1999 (TREMBLrel. 12, Created)

01-NOV-1999 (TREMBLrel. 12, Last sequence update)

01-JUN-2001 (TREMBLrel. 17, Last annotation update)

KIAA0926 PROTEIN (NALP1) (NAC-ALPHA SPLICED VARIANT).

KIAA0926 OR NAC.

Homo sapiens (Human).

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

NCBI_TaxID=9606;

[1]

SEQUENCE FROM N.A.

RC TISSUE=BRIN;

RC MEDLINE=99246063; PubMed=10231032;

RA Naase T., Ishikawa K., Suyama M., Kikuno R., Hirose M., Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;

RT "Prediction of the coding sequences of unidentified human genes. XIII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro.";

RL DNA Res. 6:63-70(1999).

[2]

SEQUENCE FROM N.A.

RA Martinon F., Hofmann K., Tschopp J.;

RT "NALP1 a novel NACHT, LRR and PYD containing protein.";

RL Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.

[3]

```

RP SEQUENCE FROM N.A.
RA Chu Z.-T., Xie Z., Godzik A., Reed J.C.;
RT "NAC: an Apaf-1/Ced-4 family member regulates the cytochrome c pathway for apoptosis.";
RL Submitted (Jan-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AB023143; BAA76770.1; -
DR EMBL: AF310105; AAC30288.1; -
DR EMBL: AF229059; AAK00748.1; -
DR InterPro: IPR000767; Disease_resist.
DR InterPro: IPR001611; LRR.
DR InterPro: IPR003592; LRR_out.
DR Pfam: PF00560; LRR; 2.
DR PRINTS: PR00364; DISEASERIST.
DR SMART: SM00370; LRR; 5.
SQ SEQUENCE 1429 AA; 160944 MW; 6C5CB8EFD2819435 CRC64;

```

Query Match 4.0%; Score 212; DB 4; Length 1429;

Best Local Similarity 20.8%; Pred. No. 1.7e-05;

Matches 184; Conservative 97; Mismatches 290; Indels 314; Gaps 36;

```

OY 142 KDQHHREVLTLNGLLQALOSP--CITEGSGKSTLQRIAMWGSK----- 190
DB 305 EENRGHLIEIRDLFGSLDQEPRIYIIGAAGIKSTLARQVKEAMRGOLYGRFOHV 364
OY 191 ---CKALTFRKFFVFLRLSRAQGLFETLDQDLIDPGTIRKQTFMMLKLRQVLEL 246
DB 365 FYFSRELAQSKVSLAELIGDG-----TATPARPQ-----ILSRPERLFI 408
OY 247 LDGYNEFRPQNCPEIEALIKENHFRKNMVIYTTTECLRHIFQFGALTAEGVDMTEDSAQ 306
DB 409 LDGVDE-----PGWVLDQPSSELCIHMNSQ-----POPAD 437
OY 307 ALIREVLIKE-LAEGLLQIQSKRLNLMKTPYFVITCALIQMGSEFPHSHTOTTFHT 365
DB 438 ALLGSLTGLTKTLPEASFLITARTALQNL--PSLEQARWVEVLGFSE--SSREYEFYR 493
OY 366 FVD-----LTIQKKKKKKGAASDFIRSL--- 390
DB 494 FTDEROAIARFLVKSNEKELMALCLVPWVSWLACTCLMOQMKREKLTSTTTTLCIH 553
OY 391 -----DHGDLALGVESHKRFDELQVSSVNE-----VLTITGLICK 429
DB 554 YLAQALQAPLGPOLRDLCL-SLAAGIWKKTLPSPDDLKRGHGLGAIITFLKMGIIQE 612
OY 430 YTAQRFKPKYKFFKSFQETAGRLSLTSHPEEYTKNGYLQKWSISDITSTYSS 489
DB 613 HP---IPLSYFIHLCFOEPFAA--MSYVL-----EDEKGRG----- 644
OY 490 LIRYCGSSVEATRAVMKHLAAVYOHGCLLSIAK--RPLMRQESLOSVKNTPEOELIK 547
DB 645 -KHSNCIIDELEKT-----LEAYGIHG--LFGASTTRFLGLISDGEREMENIFHCRLSQ 696
OY 548 AININSFEVC-----GIHLQESTKSALSOEFAFQOKSLYINSGN---IPD 593
DB 697 GRNLQWQWPSIQLLQPHSLHCLYETRNKTFLOVA--AHFEEMGCVETDMELVCT 755
OY 594 YLDFEFELPNCASALDFIKLDFYGGAMASWEKAEDTGIMBEAPETYPISRAVSIF- 652
DB 756 FCIKFSRHVK-----KIQLEIG-----RQHRSTWSPTMVY--LFR 788
OY 653 -----FMWKEFRTLEVTLRDFSKLNKODITYLGIKIFSSATSLRLQIKRCAGVAGSLSL 706
DB 789 WVPYTDAYWQILFSLVAKYT--RLKKELD-----LSGNSLSHSA 824
OY 707 VLSTCKNIYSILMVEASPLTIEDERHITSVTNLKLTSHDLQONRLPG--GLTDSLGKLNKL 765
DB 825 VKSLCK-----TLRRPCL-----LETLRLAGGGL----- 850
OY 766 TKLIMDNIKMNEDAIKLAEGKLNKMKCLFHLTHLSDIGEMDVIKSLSEPCDLEI 825
DB 851 -----AEDCKDLAFGLRANQTLTDLSPNVLTDAKAKHLCORLPSPCKLQRL 899

```



```

Db 554 YLAQALQAPLPQRLDLC-SLAAGIWKTKLPSPDDLRRKHGLDGAISTFLKMGILQ 612
QY 430 YTAORFKPKYEFKPSOEYTAGRLSSLTSHPEEYTKNGYLQKAVSLSDITSYSS 489
Db 613 HP---IPLSIFILCHCOEFPAA--MSYVL-----EDEKGRG----- 644
QY 490 LLRYTCSSVEATRAVMKHLAAVYOHGCLGLSLAK--RPLMRQESLQSVKNTTEOELK 547
Db 645 -KHSNCITIDEKT-----LEAYGIHG-LPGASTTRFLGLLSDGEGEMENIFHCRLSQ 696
QY 548 AININSFEVC-----GHLVQESTSKSALSQEFEEAFQOKSLYINSGN---IPD 593
Db 697 GRNLMQWVPSLQDLLQPHSLSLHCLYETRNKTFELQVM-AHFEEMGMCVETDELLVCT 755
QY 594 YLDFEFELPNCASALDFIKIDFYGAMASWEKAAEDTGGIHMEAPETYIPSRAYSLF- 652
Db 756 FCIFSRHVK-----KQILIEG-----ROHNSWSPYVW-LFR 788
QY 653 -----FNMKQEFRTLEVTLRDFSKLNKQDITYLGKIFSSATSLRLQIKRAGVAGSL 706
Db 789 WVPYTDAYWQILFSVLKVT-RNLKELD-----LSGNSLSHSA 824
QY 707 VLSCKKNYSILMVEASPLTIEDERHITSVNLKTLSHDLONQRLPG-GLTDSIGNLKNL 765
Db 825 VKSLCK-----TLRRPRCL-----LETRLACGGLT----- 850
QY 766 TKLMDNKNNEEDAIKLAEGIKNLKKMCLFHLHLSDIGMDIYKSSSEPDLLEI 825
Db 851 -----AEDCKDLAFGLRANQTLTELDLSFNVLTDGAKHLCORLROPSCKLQRL 899
QY 826 QLVSCCLSANAVKILIAONLHNLVKLSILDSENYLEKDGNEALHE 870
Db 900 QLVSCGLTSDCCQDLASVLSASPCLKELDLQONNLDDVGVALICE 944

RESULT 15
Q9BZ28 PRELIMINARY; PRT; 1443 AA.
AC Q9BZ28:
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE NAC-DELTA SPLICER VARIANT.
GN NAC.
OS Homo sapiens (Human).
OT Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiinae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Chu Z.-L., Xie Z., Godzik A., Reed J.C.;
RT "NAC: an Araf-1/Ced-4 family member regulates the cytochrome c pathway
for apoptosis."
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF229062; AAK00751.1;
SQ SEQUENCE 1443 AA; 162237 MW; C30EB9BE9EC82FE96 CRC64;

Query Match 3.98; Score 207.5; DB 4; Length 1443;
Best Local Similarity 21.0%; Pred. No. 3.4e-05;
Matches 173; Conservative 91; Mismatches 280; Indels 281; Gaps 33;

QY 142 KDQHHNVEQTLGLGLQALQSP--CIIEGSGKSGKSTLQRIAMLMSGK----- 190
Db 305 EENRGHLEITDLFGPLDQEPRIYILQGAAGIGKSTLARQKAVWGQGLYGDROHV 364
QY 191 -----CKALTKRFVFFPLISAQGLFETLCDQLDIPGTRKQTFMAMLLKRLQRYFL 246
Db 365 FYFSCRELAAQKVVSLAELIGKDG-----TATPAPIRQ-----ILSRPRLIFT 408
QY 247 LDGYNERRPONCPREIALIKENHRFKNNVITTTTECLRHTRQFGALTAEVGDMTEDSAQ 306
Db 409 LDGVDE-----PCWVLOEPSSESLCIHWSQ-----POPAD 437

```

```

QY 307 ALIREVLIKE-LABGLLIQIOKSRCLRNLKPTLFVYITCAIQMGSEFHSHTQTLFHT 365
Db 438 ALDLSGLKTLTPESASFITARTALQMLI--PSLEQARWVEVLGFSE--SSREYFARY 493
QY 366 FYD-----LTIQKNKHKKGVASDFTRSL--- 390
Db 494 FTDERQAIRARIVKSNKEMALCLVPVWSVLACTCLMQQKRRKREKLLTJSTKTTTLC 553
QY 391 -----DHGDLALBGFVFSHKFDELQDVSSVND-----VLTGGLCK 429
Db 554 YLAQALQAPLPQRLDLC-SLAAGIWKTKLPSPDDLRRKHGLDGAISTFLKMGILQ 612
QY 430 YTAORFKPKYEFKPSOEYTAGRLSSLTSHPEEYTKNGYLQKAVSLSDITSYSS 489
Db 613 HP---IPLSIFILCHCOEFPAA--MSYVL-----EDEKGRG----- 644
QY 490 LLRYTCSSVEATRAVMKHLAAVYOHGCLGLSLAK--RPLMRQESLQSVKNTTEOELK 547
Db 645 -KHSNCITIDEKT-----LEAYGIHG-LPGASTTRFLGLLSDGEGEMENIFHCRLSQ 696
QY 548 AININSFEVC-----GHLVQESTSKSALSQEFEEAFQOKSLYINSGN---IPD 593
Db 697 GRNLMQWVPSLQDLLQPHSLSLHCLYETRNKTFELQVM-AHFEEMGMCVETDELLVCT 755
QY 594 YLDFEFELPNCASALDFIKIDFYGAMASWEKAAEDTGGIHMEAPETYIPSRAYSLF- 652
Db 756 FCIFSRHVK-----KQILIEG-----ROHNSWSPYVW-LFR 788
QY 653 -----FNMKQEFRTLEVTLRDFSKLNKQDITYLGKIFSSATSLRLQIKRAGVAGSL 706
Db 789 WVPYTDAYWQILFSVLKVT-RNLKELD-----LSGNSLSHSA 824
QY 707 VLSCKKNYSILMVEASPLTIEDERHITSVNLKTLSHDLONQRLPG-GLTDSIGNLKNL 765
Db 825 VKSLCK-----TLRRPRCL-----LETRLACGGLT----- 850
QY 766 TKLMDNKNNEEDAIKLAEGIKNLKKMCLFHLHLSDIGMDIYKSSSEPDLLEI 825
Db 851 -----AEDCKDLAFGLRANQTLTELDLSFNVLTDGAKHLCORLROPSCKLQRL 899
QY 826 QLVSCCLSANAVKILIAONLHNLVKLSILDSENYLEKDGNEALHE 870
Db 900 QLVSCGLTSDCCQDLASVLSASPCLKELDLQONNLDDVGVALICE 944

```

Search completed: March 25, 2002, 10:59:51
 Job time: 246 sec

THIS PAGE BLANK (USPTO)

Db 778 QRLRPYRFLGPFQGEFLAVALRLELLSSDQEDODGLGLYLRQINSPLKAMSIHTFLK 837
 Oy 493 YTCG--SSVEATRAVMKHLAAVYOHGCLLGISIAKPLMRQESLOSVKNTTEOITIKAINI 551
 Db 838 YVSSHSSNAAPVNVSHL-----LQVKEKSELNNSEMEDYMKLHP 879
 Oy 552 NS--FEVECGHILYQ-----ESTSKALSQEEFAFFOGSKLYI 586
 Db 880 EALLMTECLRGMLQOLSPESSLFISENMLRICLFNFHESNIVYACSPYILOPLRGRTDLL 939
 Oy 587 NSGNIPDYLFDFEFHLPCNASALDFIKLDFYGGAMA-----SWKA----- 627
 Db 940 KYLSL-QY---FWDH--PETLLLLKISIKISLNGNMVWQRIIDESLIEKESQVPPTIDDY 994
 Oy 628 -----AEDTSGIHMEAPETYIPSRVSLFFPMNK-----QEFRTLEYTLTD 668
 Db 995 AIAFDPINEYOKNLSEKHHIITKKEDMKHQIPLINISTY--WKLSPKPKYIKRLEQVATN 1052
 669 FSKLNKODITYLGIKISSATSRLQIKRCAGVAGSLSLVSTCK-NIYSLMVEASPLTTE 727
 1053 TGPADQALLQVLMNEVFSASQSIIEFLSDSSGFLIESIRALELSKASVYFCMSRLSESR 1112
 Oy 728 DERHITSVYTKLSTHLDLQNRPLPGITDLSGLNKNTKL----- 769
 Db 1113 DKKLLTLPLPLSLEVS--ETNQLPDLFHNLMKFLGLKELCVLRDSKPDVSLVPGEEP 1170
 Oy 770 ---MDNIKA---NEEDAIKAEGLKNLKKMCLFPL--THSDIGEGMDYIVKSLSEPC 820
 Db 1171 NLHMEKLSIRTSIESDLSKLVLKIQNSPNLHVFLLKCNFLSNC--EPLMTVLASCCK--- 1226
 Oy 821 DLIEIOLVSCCLSA-NAVKILIAQNLHNLVLSILD--SENVLEKDGNEALHELIDRMNVL 878
 Db 1227 -LREIFESGRCFEMFTVNIPL---NEVFLKILNLRDQOPFDKETSEKFAQALGSLRL 1281
 Oy 879 EQLTALMLPRGCDVQSSLSLLKHLAEVPOLVKLGKMMRLDTETRL-GAFEGKNPLK 937
 Db 1282 EK--LFVPIPGDGIHOYAKLVLROCLDPCRLVLFATLDDSDVLEIKAGATRG----- 1333
 Oy 938 NFQOL-NL---AGNRRVSSDGMVLAEMGFENLKQVFPDSTKEFLPD-----PALVRKLS 988
 Db 1334 GFQKLENLDTLTHMKITIEEGYRNFQYVLDNLPLNKNLDS--RHIPESIQIQAIVYKALG 1391
 Oy 989 QVLSKTLFQEARLVYGVQFDDDLSTVI 1015
 Db 1392 QCVSRPLSTRLRLGSLWLDDEDIKVI 1418
 2
 5478
 neuronal apoptosis inhibitory protein - human
 N:Alternate names: NAIP
 C:Species: Homo sapiens (man)
 C:Date: 05-Jan-1996 #sequence_revision 05-Jan-1996 #text_change 02-Feb-2001
 C:Accession: A55478
 R:Roy, N.; Mahadevan, M.S.; McLean, M.; Shuter, G.; Varaghi, Z.; Farhanli, R.; Baird, S.
 d, T.O.; de Jong, P.J.; Surh, L.; Ikeda, J.E.; Korneluk, R.G.; Mackenzie, A.
 Cell 80, 167-178, 1995
 A:Title: The gene for neuronal apoptosis inhibitory protein is partially deleted in indi
 A:Reference number: A55478; MUID:95112344
 A:Accession: A55478
 A:Molecule type: mRNA
 A:Residues: 1-1232 <ROY>
 A:Cross-references: GB:U19251
 C:Genetics:
 A:Gene: GDB:SMAE; SMA
 A:Cross-references: GDB:120378; OMIM:600354; OMIM:253300
 A:Map position: 5q12.2-5q13
 C:Keywords: apoptosis; ATP; glycoprotein; nucleotide binding; P-loop; transmembrane prot
 P:94-110/Domain: transmembrane #status predicted <TM1>
 P:470-477/Region: transmembrane #status predicted <TM2>
 P:479-496/Domain: transmembrane #status predicted <TM3>
 P:476/Binding site: ATP (Lys) #status predicted

Query Match	Best Local Similarity	24.1%	Pred. No. 3.8e-25;	Mismatches 226;	Conservative 155;	Mismatches 341;	Indels 215;	Gaps 33;
69	FLKSLKENMYPIFEODLNGO----	SLFPHQTSGLD-----	100					
Db	344	FLQNNKKS-SAEVTPDQSGELCELLETTS	ESNLDESINAVPIVPEMAQGEAQMFGDAKN	402				
Qy	101	LAQDLKDLVHTPELNF-----	YPLGEDIDILFNLSKSTTEPYLMKKDDHNR	148				
Db	403	LMEOFLRAATSAFSRHMSSLDLSDTLATDHLGLCDLSI--	ASKNLSKPV-----	449				
Qy	149	VQDLTLNGLLQALQSPCIIEGESGKSPYLORIAMWGSGKALTKREKVFPLRS--	206					
Db	450	QEPVLVPEFGNLSNMCVEGEGAGSGKTVLKLIAFLAMSGCCPLLRFPVLVLSIST	509					
Qy	207	RAQGGFLFEYTCOLDLDIPCTIRKQTFMAMLLKRLQVLEFLLDGYNEF--	KPQNCPEIEAL	264				
Db	510	RDDEGLASTICDQLLEKESGVTETMCKRNIIQQLKNQVLEFLDDYKEICISIPQ--	YIGKL	566				
Qy	265	IKENRFRKMMVYVTTTTECLRIHQGALTAEXGDMTEDSAQLIBEVLIKELA--	BGLL	322				
Db	567	IQKNHLSRCLLAIAVNTNRARDIRYLETILEAQEPFYTVLIRKLFSHNRTRLKKFM	626					
Qy	323	LQIQSRCLRNKMPKPLFVAVITCAIQMGSESEPHSHQTFLEHFHYDILLQKNHKHKGVA	382					
Db	627	VYFGKNQSIQKIQKTPLFVAALCAHMFQYRPDPSPFDVAVAFKSYMELSLRNK-----	A	680				
Qy	383	ASDFIR-SLDHCGDLALLEGVFSHKFDFELQDV--	SSVNEDVLLTGLGLKYTAQREPKY	439				
Db	681	TEIILKATVYSSCGELAKLKFSCCFEFNDDILAEGADEDEDLTMCMSKFTAQRLRPFY	740					
Qy	440	KRFHKSFOGYTAGRRLLSLLTSHPEEYVTKGNQYLQKMSVLSIDITSYSSLLRYTCSSV	499					
Db	741	REFLSPAFQFELGLMRILIELDSDRQEHODGLYHLKQINSPPMTVSAYNNFLNV--	SSL	798				
Qy	500	EATRA---VMKHLAAYQHGCLGLSLIAKRRPLMKROESLQSVKN-----	TTEQEI	545				
Db	799	PSTKAGPKIVSHLLHLYDN-----	KESLENISENDYKLHQPEISLQML	843				
Qy	546	LKAI-----NINSVVEGIIH-----	YQESTSKALSGEFEPFGKSLYINSGNIPD	593				
Db	844	LGLMWOICPOAFYSNVSSEHLVIALKTAQOSNT--	VAACSFVYLQFLOGRLLTGALNL-Q	901				
Qy	594	YTFD-----	FFEHLPNC--ASALDFIKLDFYGG--	AMASWE	625			
Db	902	YTFDHEESLSLRSIHFSIRGNKTSIPRAHRSVLETCDKQOVPTIIDQDVASAFENMWE	961					
Qy	626	-----	KAEDTGGIHMEADETY-IPSRAYSLEPNNKOEFTLEVTL	666				
Db	962	RNLAEKEDNVKSYMQRORASPDLSGTGYWLSKRYOKIP-----	CLEVV	1006				
Qy	667	RDESKLNKODITYLKIFESATSLRLQIKKACAVAGLSVLSTCK--	NITYSLMVEASPLT	725				
Db	1007	NDIDVYGQMLELMTVFSASQRIELHLNRSKFIESIRALELSKASVYKCSISKLELS	1066					
Qy	726	IEDERHITSVNTKLTSIH--	DLQNRQLPGGLTDSIGNLKNITKLIMDKM-----	775				
Db	1067	AAEQELLTLPLPSESLEVSSTIQSDQIPFN--	LDKRLCLKELSVDEGNINVSVPREE	1124				
Qy	776	-----	NEEDAIKLABGLNKKKCLPHLPHL-----	SDIGESMDYIVASLS	816			
Db	1125	PNFHHMERLLIQISAEYDPSKLVASLPNTISKLINLEGOQPPDETSSEKFAVILIGLS	1184					
Qy	817	SEPCDLEFIQVASCASANAVKLIILQNLHLVKLSII	853					
Db	1185	-----	NIEELLPLPGDGITRYAKLIIQCCQGLHCLRVL	1217				


```

0Y 378 HGVAA5DFRSDH-----CGDLAGEVSHKDFELOVSSV-----NED-- 419
Db 627 LRGVIALSF-KSLDEEKKIFLDIACFL-----KMDITKEEYDILKGGCLMAEAA 677
0Y 420--VLTTLGLCKYTAORFKPKRYKFEHKSFOEYTAGRRLSLLTSHEP-----EYU 468
Db 678 LRVLIQKSLTILTDPL-----WMHQIRD--MGRQVHKHSSDPEKRSRLMDRGEIM 730
0Y 469 KGNGLYQKNWISIDISTYSSSLRYTCGSSVEATRAVKKHIAAVYQHCGLLGISTAKRPL 528
Db 731 NVLDYKKGTSISIGIYDFN-----KKFARDHTADEIFSSNLNNP-- 771
0Y 529 WROESLOSXKNTTEOEELIKAININSFECCGILHXOESTSKASLODEFEAFQGSXYINS 588
0Y 772 ---GIVSFNYYKNNKIVR-----FPAEKKRSEITTPVESFAPKKIRLQ 815
0Y 589 GNIPDYLPEFEEHLPNCASALDFIKLDFYGGAMAWEKAABETGTGIMEABETYPISR- 647
Db 816 INNVELEGD-LKLLP--SELKWIO-----WK-----GPLENLPPDLSRLQ 854
0Y 648 -AVSLEFNMKOEPRTE-----VTLR-----DFSKLNKODITYLGIKTFSSA 687
Db 855 GYVLDSSESGVRKVTLPKRKEDENILKYVNLRGCHOLEAIPDLSNNA-----LEKLVLR 909
0Y 688 TSLRIQKRCAGVAGS-TSLVSTCKNYSLWVNASPLTIDBRHTSVTNKTL----- 741
Db 910 CNLVVPPSVGNLQKLLQDLDRCCSSLSFEFLGDVSGLKCLEKFEFLSGCSNLSVLPENIG 969
0Y 742 SJHLDONORLPG---GLTDSIGNIKNLTKLIMDNIKMNE-----DAI 781
Db 970 SMPCKLELLDGAHAISSLPSIFRLOKREKLSLMGCRSIEELPSCGYLTSLEDLYDPT 1029
0Y 782 KLAE-----GLKNLKMCLPHLTHLSDIGECMDYI--VKSLSSEPCDLEBQL---VS 829
Db 1030 ALRNLPSISIGDLNKLQKHLMRCTSLSTIPETINKLSKELFINGSAAVEELPIENGSL 1089
0Y 830 CC--LSANAVKILAO-----NLNVLVSLTLDISENYLEKNGEALHELIDRMANVLEO- 880
Db 1090 CLTDLNAGDCKFLKQVYSSSIGLNSLQ-----LDSTPIEALPEIGDLHITROL 1141
0Y 881 -----LTALMLPWG-CDVQSSLSLKLHEVPO-----LYKLGLKNRRLTDEIR 925
Db 1142 DLNRCKSLKALPRTIGKMDLYLSLNLVGSNIIEELPEFGKLEINIVELRNNCKM----LK 1197
0Y 926 IIGAFRG--KNPLKNPOLNLAGNRVSSDGNLAFGVFENLKOLVF--FDSTKEFLPDP 961
Db 1198 RLPSKSGDLSLRLXYQOETLYAELPESFGNLSNLMVLEMLKPPRISRISSENVGTSEEP 1257
0Y 982 ALVRLQSVYSLKTLFQEARLVGQOF---DDLSVITGAPK 1021
Db 1258 RYV-PVPSNFSKLLKLEELDACSWRISGKIPDDLEKLSLMLK 1299

```

RESULT 5

hypothetical protein T18B16.20 - Arabidopsis thaliana
 A:Alternate names: hypothetical protein F13C5.220
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 30-Apr-1999
 C:Accession: T04426; T05042
 R:ByRef: M.; Benes, V.; Rechmann, S.; Borkova, D.; Ansonge, W.; Bancroft, I.; Mewes, H.W.
 submitted to the Protein Sequence Database, April 1998
 A:Reference number: Z15359
 A:Accession: T04426
 A:Molecule type: DNA
 A:Residues: 1-1405 <BEV>
 A:Cross-references: EMBL:AL021687
 A:Experimental source: cultivar Columbia; BAC clone T18B16
 R:ByRef: M.; Pohl, T.; Weisengger, T.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Schuel
 submitted to the Protein Sequence Database, February 1998
 A:Reference number: Z15395
 A:Accession: T05042
 A:Molecule type: DNA

A:Residues: 1-1405 <BEM>
A:Cross-references: EMBL:AL021711
A:Experimental source: cultivar Columbia; BAC clone F13C5
C:Genetics:
A:Map position: 4
A:Introns: 893/3, 1154/2, 1183/3, 1192/3, 1198/2, 1214/3, 1251/1, 1282/2, 1327/1, 1381/2
A:Note: T18B16,20; F13C5,220

Query Match	3.5%;	Score 187;	DB 2;	Length 1405;
Best Local Similarity	20.1%;	Pred. No. 0.0056;		
Matches 201;	Conservative 151;	Mismatches 325;	Indels 322;	Gaps 51;

QY	163	SPECIEGSGKSTLORIAMIMGSGCKALFKRF-VFELIRISA---OGGLFETLD	21
Db	27	SKYLVGAGGKGLWLAKVES-----GRVQEKVNLMLHNKKIDIEKSLYELIAA	78
QY	219	QL-----LDIP-GTIRKQTFMAMLIKLRQVLFLLDGYNEFKPONCP----	259
Db	79	QLSTIIEEEDDEEDLDYPLESLEKIKEBMKKKKDNLLILDEGSMTEDEYMOEL	138
QY	260	EIEALIKENHFKMMVYVTTTECLRHIRQGLTAVGSMTEDSQAOLIRE-----	311
Db	139	NLODFLEKYSVK--ILVTRDPE--REKE--STTIKVGPLTKESJDLHDMEDLLTSE	192
QY	312	-----VLI-----KEIABGLLQLOKSRCLRNLMKPTPEFVITCAIOWGESEFHSHTQ	359
Db	193	TSEDMPVLKRLCDNKEIKETPLMS-----CILSKSGLPAAIVYLKLSINIKSMSAKO	247
QY	360	TTLEHTF-----YDLLOKNNKHKHGVASDPTIRSL	390
Db	248	RKIKERELLSSLDEAASAKNAIDRSRYNPVLOLSTELKLPDEYKRPVIAIC--FWHIL	305
QY	391	D-----HCGDL-----ALEGVFS--HKRFDELQDVSVNEDVLLTGLLCKYTAQRFK	436
Db	396	DFKYSGCAYYRDLIVHMLEGYDPRKSVDAKVOBESHILMD--FNNRGILKIQEDNMVY	364
QY	437	PKYK-----FFHKs-----QOETTAGRLSSLTSHPEPVTKKNGSYLOKMSID	482
Db	365	PEFSMNLLDLDGCGFFGSRSLGDRVYGG-----DKRRGLG--KIILIDD	408
QY	483	ITSTYKS-----LLEYTCGSSVEATRAWKHIAVYQ---HGCILGJSIAK	525
Db	409	MIQLOSKKNMITTILNSGNLRREYHGKRFEPKEMQDELVVLLFEPTFHELVLSTSKK	468
QY	526	RPLMRQESLOSVAKNTTDOEILKAININSYECGIIHLQOESTSKSAISOEFAEFQOKSLY	585
Db	469	K-----LREVLIRCDLDINDIKLSGLO-GILHVEVSGASSLV-----	505
QY	586	INSGNIPYLRDFPEHLRNCAASALDFIKLDFYGGAMSWEKAEDBGCI-----HW	636
Db	506	-----NIPD-----DFKNNMQLOS-----LNSGLAISSPSTIEKLSMLCFLIRHCELS	552
QY	637	BEAPETYIPRSRAVSL-----FFNNKOEERTEVTLRDFSKLNKODITYLAKIES	685
Db	553	QDLPRFIVETKRLEVIDIHGARKLESYFEDRWKDKYKKNKNPAOL--QLTEHLDD--ES	608
QY	666	SATSIRLOIKKACAGVAGSLs-----LVISTCKNITSYLVAVESPILT--IDEBRHTSY	735
Db	609	EKTIRLPIFLKLDSTNDFSTMPILTRLLLRNCRKRL--POLRPLNIQIIDACGATDL	667
QY	736	TNL-----KTLSDHLOMORLPGGJLTSLGKMLTKLIMDNIKMNEDEAIKLAEGJ	787
Db	668	VEMLEVCLEEKKELRIDMKTSLP--ELADTIADVNLKLLLRNCLTIE-----L	718
QY	788	KNLKMCLFHLTHLSDIGEMDYTVKSLSEPCDLEBIOVSCCLSANAVKLLAONLHNF	847
Db	719	PSIEK-----LTHL-----EVPVSGCIK--LKNINGSFGEM	748
QY	848	VKLSTLDSBNYLEKODGEALHELIDMKNVLEQLTAMLPWCGDVGOSLSSLLKHLHEVP	907
Db	749	SYLHEVNLSTENLS-----ELPDKISTELSNKELTIL-----RKCSKLTLE	789

Db 708 PENMASLVEFLNKGCTGLESJPKINLRSLKTLILSNCSNL-----EEFWVIS 754
 QY 735 VNNKLTSLIHQONRRLGGGLTDSGNLKNLKLMDINKMNEEDAIKLAGLKKMKM- 793
 Db 755 EFL-LTTLTLDGTATKTLTP--QDMVKLTSLVKLTKMDCM---LVKLPEEFDLKLVQ 805
 QY 794 -----CLFHLTHLSDIGEMDYI-----VKSLS-EPDDEEIOLVSCSLA 834
 Db 806 ELVCSGCK-RLLSLPDVKNKQCLDILDLGATIKIPHISLEERLCLSRNEKISCL--S 862
 QY 835 NAVKTLA0---NLHNLVKL-SIIDLSENYLEKGN-----BALH-- 869
 Db 863 NDIRLSOLKMLDKYCTKLVISIPELPTNLQCLDANGCESLTTVANPLATHLPBQIHST 922
 QY 870 -----ELIDRNN-----VLEQLTALMLPWGCDVQGSLSLKLHLEVPQVLKLG-LKMKWR 918
 Db 923 FLFTNCDKLDRAKGFEALFSTCFP-GCEV---PSWFCB-EAVGSVTLKLLPLHW- 975
 QY 919 LDTFIRILG 928
 Db 976 ---NENREVG 982

RESULT 9
 T06403
 resistance complex protein I2C-1 - tomato
 C:Species: Lycopersicon esculentum (tomato)
 C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 17-Nov-2000
 C:Accession: T06403
 R:Ori, N.; Eshed, Y.; Paran, I.; Presting, G.; Aviv, D.; Tanksley, S.; Zamir, D.; Fluhr,
 Plant Cell 9, 521-532, 1997
 A:Title: The I2C family from the wilt disease resistance locus I2 belongs to the nucleot
 A:Reference number: 215652; MUID:97290204
 A:Accession: T06403
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1220 <ORF>
 A:Cross-references: EMBL:AF004878; NID:g2258314; PIDD:AA863274.1; PID:g2258315
 C:Genetics:
 A:Gene: I2C-1
 A:Map position: 11
 C:Function:
 A:Description: confers resistance against Fusarium oxysporum
 C:Superfamily: disease resistance protein RPS2; leucine-rich alpha-2-glycoprotein repeat

Query Match 3.28; Score 170.5; DB 2; Length 1220;
 Best Local Similarity 19.8%; Pred. No. 0.043;
 Matches 231; Conservative 161; Mismatches 393; Indels 383; Gaps 61;
 QY 8 SRALQRMGMVYIKQITDDLEFVWVNLNREVNIIICEVEVDDAAGITIMILKSGSECN 67
 Db 241 TKGLLQETGISTDLK--ADD-----NLNOLQVLRKADNMLNQ-----LOVKLKEKLN 284
 QY 68 -LFLSLKE-W--NYPLFQDLNGSLFQTSSEGLDD-----LAODLKDYH 110
 Db 285 GKRFVLVDVDMWNNYFPMWDL--RNLFLQ---GDIGSKIIYVTRKESVALMDSGAIY- 338
 QY 111 TPSPFNFPVIG---EDIDIIENLKSFTFEPVLMRKD-QHHHREQL-----TLNGL- 157
 Db 339 -----MGLSSEDSMALEFKRHS-----LEHKDREKHEPEFEVQKQADCKGLPL 383
 QY 158 -LQALQSCITIEGESGKSKSLRLQRIAMLNGSGKALKTRKFVFFLRSLAAGLFFETL 216
 Db 384 ALKALAGLRKSKSEVDENRILREI---WELPSC-----SNQILPAL 423
 QY 217 CDQLIDIGTIRKOTFMAMLLKRLQRLVLLDGYNEFKPQNCPEIEALIKENHRP-KMY 275
 Db 424 MLSTVDLPAHLKQ-----CFATCAIYPPDYOPRKQY 455
 QY 276 IVTTTECLRIHROGA-----LTAEGVDMTEDSAQALIREVLLKELAGLLLOIQK 327

Db 456 IHLWIANGLVH--OFHSGNYFIELRSKLEFEMASEPSEPDVEEFLMHLVDNLAQIASS 513
 QY 328 SRCRLNLMKTLFLVYITAI-----QMGSE-----FHSITQTL-----EFTFDL 370
 Db 514 NHCIRLEDNKGSHMLEOCRHMSYSIGODEFEKLSLEFSQRLTLPIDIQH-YSKKL 572
 QY 371 IQKNKHKHGVAAASDFITSLDHC-----GDALGEGVSHFDELODVSSVNDVY- 421
 Db 573 SKRVLHNLPTLRSLRALSLSHQYLEVLPNDLFKIKLRRLDLSERSITKLPDSIVLY 632
 QY 422 -LTGGLL--CYYTQREPKYKFFHKSFOEYTAGRRSS--LTSHPDEYTKGNGYL-- 474
 Db 633 NLETLSSCYLELPLQMEKLIINRLHDSNRRLKMPHLRLSLQVLYGAKFLVG 692
 QY 475 -OKWVSIDITSTYSSILRYTCGSSV---EATRAVMKHLAAYVGHGCLLGSIKKRPLMR 530
 Db 693 GWRMEYELAEANLYGSLISLELEVNDREAVKAMKRNHVEQ-----LSE---W- 741
 QY 531 QESLOSVKNTFEQEL-----KAININSFVECGIHLVQESTSKSALSQFEAFEG 581
 Db 742 SESISADNSOTEROLDLDELREKHNKAVEITGY----- 774
 QY 582 KSLYINSGNIDPLYLD-----FEEHLPCAS-----ALDFIKL----- 614
 Db 775 -----RGTFPNWVADPLFVKLVHLILRNCKDCXSLPALQOLPCLFELSGMHGIRVYT 829
 QY 615 -DFYGGAMASWEKAEDTGGIHMEAPETTYIPSAVSLFFWKKQ-----EFTLEVTL 666
 Db 830 EEFYG--RLSKKPFNSLVKRFEDMP-----WKQWHTLGIGFPTLE--- 871
 QY 667 RDFSILNKQDITLYGKIPSSATSLRLQRCAGVAGSLVLT-----CKNIYSIM 718
 Db 872 -----KLSIKNCPLESLERPIQFSSLSKLRLDIDCKSVTSFP 907
 QY 719 VEASPLTIE-----DERHITSVNLKTLSTHQLONRPLPGLTDSGLNT 762
 Db 908 FSILPTLRKIRKIGSCPRLKLEAPVGEVFEYLVIDGCVDDISPEFLPTARQLSIENC 967
 QY 763 KNLKRLMDINKMNEEDAIKLAGLNLKMKCL-----FHLTHSDIGEGMDIYVKSLS 817
 Db 968 HNVTRFLPTATESLH-----IRNCEKLSMAAGAAQTLTSLNIG-----CKRLKC 1013
 QY 818 EP-----CDLEETIOLVSCCSANAVKILQNLHNLVKLSTIDSEVYLRK--DGNALHELI 872
 Db 1014 LPELLPSLAKELRYTC-----PEIGELRPN---LQILDI--RYCKKLVNKGKEMH--- 1059
 QY 873 DRMNVEQLTALMLPW-GCDVQGSLSLKLHLE--EVPQVLKGLKMKWRLTDTFIRILGA 929
 Db 1060 -----LQRLTELWIKHDGSD-----EHIEHMERLPSIO-----RLFTFNLKTLSS 1099
 QY 930 FFGKNPLKNPQQLNLAGN--RVSSDGMALPFGVFNENLKOLVFPFSTKEFLPPDALYRKL 987
 Db 1100 QHLKS-LTSLQFLRTIVGSLQFQSGOLSSFSHLTSLQTLQIWNFLNLQSLPESALPSSL 1158
 QY 988 SQ-VLSKTLFQEARLVGWGCFDDDDLSV 1014
 Db 1159 SHLIISNCPNLQSLPLKGMPSLSLSTLSI 1186

RESULT 10
 T15864
 hypothetical protein C56E6.6 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
 R:Fullon, L.
 submitted to the EMBL Data Library, November 1995
 A:Description: The sequence of C. elegans cosmid C56E6.
 A:Reference number: S69019
 A:Accession: T15864
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1066 <FULL>

A:Cross-references: EMBL:U39996; NID:g1055114; PID:g1055120; PIDN:AAA81094.1; CESP:C56E6
 C:Genetics:
 A:Gene: CESP:C56E6.6
 A:introns: 62/3; 110/3; 376/3; 432/3; 482/3; 574/2; 649/2; 679/2; 769/2; 833/3; 892/3; 9

Query Match 3.2%; Score 170; DB 2; Length 1066;
 Best Local Similarity 20.9%; Pred. No. 0.038;
 Matches 160; Conservative 114; Mismatches 299; Indels 194; Gaps 35;

QY 314 IKELAGLLDQKRCNLNMTPL-FVITCAIQMGSEFHSHTQTL-----FHT 365
 DB IETIQMSPQPTLKLVLNRLNLTQIPVALNDLKTQSIDLGNITILMDTNEVTFES 145
 QY 366 FVDLLQKKHKKHGAASDFIRSLDHCDLAEVFSHKF-----DFELQDVSSV 416
 DB 146 EKKVILRNK-----VARLD-----KHSFRFKIRLIDISYNOIQTV 183
 QY 417 NEDVLLTGG---LCKYTAQRFKPKPKPKHSFOEYTAGRRLSLTSHPEEVTKMG 472
 DB 184 EDSSEFTVGHMOSIDLSYNRIATLPKGM-L-KNFATLTKLKLAEWMIHATPEALDLRN 241
 QY 473 YIQKMSIDITVSSSLRYTCGSSVEATRA--VMKHLAAYQHCCLG-----LSTA 524
 DB 242 LTHLNKGNKLRIDGDLGCTDTLVELFPIANNYLEHI---PHCVLSGMKOLEHLDIS 297
 QY 525 KRELMBROESLOSAYKNTTEDEILKAININSEFEGCIIHLYOESTSKASLOEFAFPOGSL 584
 DB 298 KKNIMSLKPKTSLSTTKETSTVRRLN-----LAGNRI 331
 QY 585 YINSGIPIPLPFEPFHLPCASALDFIKIDPYGAMAS---WEKAEDTGGIIMEBAP 641
 DB 332 ---NNMSDYL--IFEHMP---LITYVDSFNRIRFISPRVEK-LKNLESFLQNNOL 380
 QY 642 TYTPS-----RAVSLFEMWKEPRTLEVTLNDFSKL-----KKODITLYGKIESSA 687
 DB 381 AHPSLFLRDLKLRLMDNNOIOKIDNF--SLADLRKLOHLSLAGNOLDI-ITENMGSS 437
 QY 688 TSLRLQIKRCAGVAGSLSLVSTCKNIYSLM-----VEASPLTIEDERHITSVT--NLK 739
 DB 438 SSELK-----SLNLAHNKIHSSRSFSDLDNLQDLRLSHNNIRITSMFSSNLK 488
 QY 740 TLSIHLOQRPLPGGLTDSIGNKNTKLMIDIKANE--EADIKLAEGKLNKKMCLFH 797
 DB 489 NLTYYDLSHNRITKILPSALYQDLPALDYLRDLNHNNEIDRDAFRSPDSQSLK----- 542
 QY 798 LTH-----LSDIGEGMDIYVKSLSSEPCDLEIOLV-----SCCLSANAV-KI 839
 DB 543 LSHNARRRSCFELGISQ-----VHQLDSSNQINETIDIFCIARGLRKLSTLANSVEKI 597
 QY 840 LAQNLNLVKLSTLIDSENLYLEKQNEALHELIDRMNV-----LEQDTALMLPMGC 890
 DB 598 NRRLLDQATFELTIDISHNGLIDVSDAFCECKLSHIKISHNYIRLMKGTVCIVCIS 657
 QY 891 DVQGLSSLLKHEEVPOLYKLGKMMRLTDTETIRILGAFGKNPLKNCQQLAGNRVS 950
 DB 658 HLTTFCEFTKEHLERTISFYIIVDSQSLTS-----FG-----NSTLSFANKVD 703
 QY 951 S--DGLAFMGVFENLKOLVFPESTKEFLP--DPALVRKLISOVLK 994
 DB 704 STED-----GAFENLSLKIDLSNNPVTSMSPTAFRDLSHSISSI 744

RESULT 11

F84547

Probable disease resistance protein (imported) - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C:Accession: F84547
 C:Accession: F84547

R:Lin: X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffatt, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.;
 Neus, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 402, 761-768, 1999

A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: AB4420; MUID:20083487

A:Accession: F84547

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1110 <STO>

A:Cross-references: GB:AE002093; NID:g6598362; PIDN:AAF18600.1; GSPDB:GN00139

A:Genetics:

A:Gene: At2g17060

A:Map position: 2

Query Match 3.2%; Score 169.5; DB 2; Length 1110;
 Best Local Similarity 18.7%; Pred. No. 0.043;
 Matches 187; Conservative 137; Mismatches 279; Indels 397; Gaps 46;

QY 150 EQLTLNGLQALQS--PCTIEGE-----SKGKSTLLO-----RTAMLG 187
 DB 67 EQLLNQIVGEQVRVIRKISQEGDEKKNMVCSTNSTGSSFIPOKRNWVDPENQIEVLG 126
 QY 188 SCGRALTKFKFVFLRLSRAOGGLEPFLCDLIDIPGTRKQTFAMLLKRLQRYFL 247
 DB 127 SQRLKELKE-----KDLISRK-----ETRIYGLMGPG-IGKTTLYKRL----- 164
 QY 248 DGYNEFPQNCPEIEALIKENRPFKNMVIYTTTECLRH-----RQFGA----- 292
 DB 165 ---YDEMK-----HNFO-----RHLMWNIROKSEKVEGTHSLERN 196
 QY 293 ---LTAEGMDTDSQAOLIREVLKELAGLLDQKRCNLNMTPL-FVITCAIQMGSEFHSHTQTL-----FHT 365
 DB 197 IKELSLDITVNDITTEMTYASVAKDELKRVLLVLDVSSKQIQGLNLMWIRGSHI 256
 QY 342 VITCAIQMGSEF-HSHTQTL-----FHT 365
 DB 257 VITTRKISISQEGYVYVPRINTDGLKQFSFYAEEDHNCYPGNLMDSTKFDYANG 316
 QY 369 --LTIQKMKHKKHGAASDFIRSLDHCDLAEVFSHKF-----DFELQDVSSV 416
 DB 317 NPLALKILGRELISIDKQDMPKRLDPLQIPYIDDLRASVD--DLSQOKEVELV 372
 QY 424 TGLLKYTAQRFKPKPKPKPKHSFOEYTAGRRLSLTSHPEEVTKMG 472
 DB 373 VA-----WFGSGDEY-----YIRSLVDTEDPOSADD-----AASEV 404
 QY 484 TSTYSLRYTGGSYEAARAVKHLAAYOHGCLLGLS---IAKRLMBROEST----- 534
 DB 405 RDPAGNLL-----ISISGRLEHMDLMTFAKKLSSLSNENNYGYQIMNHESFNAAK 459
 QY 535 -----QSVKNTTEDEILKAININSEFEGCIIHLYOESTSKASLOEFAFPOGSL 584
 DB 460 NKRMRYVNOPRRKRVTESEMDNVGI-----LIDVSEMNNMTLDKRFSEMCNERY 510
 QY 586 INSGNIPDVLDFEHLPCASALDF-IKLDYFGAMASMEKAEDTGGIIMEBAPET 641
 DB 511 LKAYN-----SQCSDCVGCKLTFPDGKCS-----HENVYLYW 546
 QY 644 --IPSAVSLFFN-----WQEFRTLEVTLNDFSKLNKODITLYGKIESSA 687
 DB 547 LQPLKRLSKAFPKMLIELNLPYKTRLMKES-----KEISKLMWVDSHSSEL 598
 QY 680 -LGRITSSATSLRLQIKRC-----AGVAGLS-----L 706
 DB 599 DISGLIGAHNIRRLNLEGIEELKTLPEQMESESLIYLNAGCTRLVSLPEFKLSKL 658
 QY 707 VISTCKN-----LYSLMVEASPLTIEDERHI--TSVNTLKTSLHLOQRPLPGGLTDS 760
 DB 659 ILSHCNFEQFPVYSCLLALYQGTALCIPSTIENLKLILLDKDEVLSLPDCIG 718
 QY 761 NLKNTKLIM---DNKIMNEDAIKLAEGKLNKKMCLFHLTHLSDIGG-----MDYIV 812
 DB 719 NURSLGELILSGSKAKFPPE-----LKETMKSIKILL-----DGTAIKQPIILL 764
 QY 813 KSLSE-----PCDLDEI-----QLVSCCLSANAVKILANLNLVKSITLDSF-- 857

Db 765 OCIOGSHSVANKTLPSNSLDYLLPSSLLSLCISGNDIESLHANISQYLHLKWLDKNCK 824
QY 858 -----NYLEKGNENALHEL----- 871
Db 825 KIKSVSLVPLNKLCKDAIGCCSLIEVGSPLAVLWTKIHCTYIFTNCKIKLDQVAESNII 884
QY 872 -----IDRMN---VLEQLTALMLPWGCDVQGS 895
Db 885 SFTWRKSQMSDALNRKNGFVLESIVSTCFP-GCEVPAAS 923

RESULT 12

T10504

disease resistance protein Cf-2.1 - currant tomato

C:Species: Lycopersicon pimpinellifolium (currant tomato)

C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 17-Nov-2000

C:Accession: T10504; T10515

R: Dixon, M.S.; Jones, D.A.; Keddie, J.S.; Thomas, C.M.; Harrison, K.; Jones, J.D.

201 84, 451-459, 1996

Title: The tomato Cf-2 disease resistance locus comprises two functional genes encoded

Reference number: Z17062; MUID:96190812

A:Accession: T10504

A:Status: Preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1112 <DIIX>

A:Cross-references: EMBL:U42444; NID:91184074; PIDN:AC15779.1; PID:91184075

A:Experimental source: cultivar Cf 2

A:Accession: T10515

A:Status: Preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1066, 'I', 1068-1085, 'E', 1087-1110, 'R', 1112 <DI2>

A:Cross-references: EMBL:U42445; NID:91184076; PIDN:AC15780.1; PID:91184077

A:Experimental source: cultivar Cf 2

Query Match 3.1%; Score 165; DB 2; Length 1112;

Best Local Similarity 21.6%; Pred. No. 0.08; Matches 139; Conservative 81; Mismatches 199; Indels 224; Gaps 31;

QY 425 GLLCKYTKORPKPKYFKHKSFOEYTAGRSLSLTSHEPEVTKNGYLOKMWISDIT 484
Db 140 GLLAQLQIR-----IFHNQ-----LNGFIKPEI-----GYLSLTKLS-LG 175
QY 485 STYSSLYATGSSVEATRAVWKHLAAVYQHCGLLSLIARPLMKROSLSQVKTTOE 544
Db 176 INFUS-----GSLPISAVGNLNLSPFLYLYNNQLSGSI-----PEISYLRSLTELD 221
QY 545 ILKAININSFVECGIHLQESTSKALSQEEA-----FFQKSLYINSGNIPD 593
Db 222 L-----SDNALNGSTIPASLGNMNNLSFLFYGNQL-----SGSTIPE 257
QY 554 ---YI-----FDPEHLN---CASALDFIKLDF---YGAWASWEKAEDTGGIHMEEA 639
Db 258 EICYLRSLTYLIDLSENALNGSIPASLGNLNLSPFLFYGNQLSG--SIPEIG----- 308
QY 640 PETYIPSAVSLFFMKKQEFRTLEVTLRDFSKLNQDITTYLCKIFSSATSLRLQIKRCAG 699
Db 309 ---YI-----RSLNLVLGSENALN----- 324
QY 700 VAGSLSLVSTCKNITYSLMEVSPPLTIEDERHITSVTNLKTLISIHLDONORLPGLTDSL 759
Db 325 ---GSLPISAVGNLNLSPFLYLYNNQLSGSI-----ASLGNLNLSPFLYLYNNQLSGSI 379
QY 760 GNLKNTLTLMDNLIKANEEDAKLAESLKNLKKMCLFHLTHLSDIGEGMDYIVK---SL 815
Db 380 GNLNLNLSMLYLYNNQLSGSIPIASLG--NLNLSRLYLYNNQLSGSIPEIGYLSLTYIDL 438
QY 816 SSEPCD-----LEEIOLVGCCLSANAVKTLIAQLNHLVLTSLIDSEN 858
Db 439 SNNSTNGITPASPFGMSLAPLFLYENQLASS-----VPEIGYLRSLNVLNIDSEN 489
QY 859 YLEKDGNEALHELIDRMNVLEQLTALMLPWGCDVQGSLSLKLHLEVPQVLKGLKMW 918

Db 490 AL-----NCSIPASFGNLNLNLSRLN-----VNNQLSGSIPE--EELGYLRSLNVLNID-- 533
QY 919 LTDEIR-ILGAFRGKKNPLKFNQOLNLAGNRVSSD-----GWL----- 955
Db 534 LSENALNGSIPASFGN--LNNLSRLNVLNNLQSGSIPEIGYLRSLNIDLSENALNGSI 591
QY 956 -AFMGVFENLQVLFDEFTKEPLDPALVRKLSQVSKLFL 997
Db 592 PASLGNLNLNLSMLYLYNNQLSGSIPEI-----GYLSLTYL 628

RESULT 13

T06145

disease resistance protein homolog F24J7.80 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 22-Oct-1999

C:Accession: T06145

R: Bevan, M.; Vitale, D.; Lianori, R.; Argyriou, A.; De Simone, V.; Bancroft, I.; Mewe

submitted to the Protein Sequence Database, April 1999

A:Reference number: Z15493

A:Accession: T06145

A:Molecule type: DNA

A:Residues: 1-1715 <BEV>

A:Cross-references: EMBL:AL021768; GSPDB:GN00062; ATSP:F24J7.80

A:Experimental source: cultivar Columbia; BAC clone F24J7

A:Gene: ATSP:F24J7.80

A:Map position: 4

A:Introns: 4/1; 122/2; 476/3; 575/3; 1088/3; 1154/2; 1512/2; 1610/1; 1662/1

Query Match 3.1%; Score 164.5; DB 2; Length 1715;
Best Local Similarity 20.3%; Pred. No. 0.16; Matches 217; Conservative 134; Mismatches 312; Indels 407; Gaps 56;

QY 53 GIIMILKSGSECNLFKSL-----KENYPLFO--DLNGOSL-FPHQSEGLD-DLA 102
Db 76 GLTNSIVK---HC-LELKKMYPDDKVDENRNALMDADLGGVSHKSDSELVEKIV 131
QY 103 QDLKDLVHTPSFLNFPLEGEDIDIFMLKSTFTEPVLMRKDQHHRYEQLTLNGLQALQ 162
Db 132 ADVROKIDRRIRIGVY-----SRLTKIYLYLCKQ 160
QY 163 SPCLIE-----GEGKKGSTLLQRIAMMGSKKALTKKFVFLKLSR----- 207
Db 161 PGCIIRSLGIMGAGIGKTTLAR-----AAVDOLSRDFEASCFIED 201
QY 208 ---AAGLPEPTCODLDIPGTRKQTFMAMLLK-LR-QRVFLDGVNEFRPQNC- 259
Db 202 FDRPEQKGFGLLEKQL-----GVNPOVTRLSILKTLRSKRILLVLDVNR--KPLGATS 255
QY 260 ---EIEALIKENHFKKNVIVTTTECLRHIRQEGALAEAGMTEDSAQALIREVLIKE 316
Db 256 FLCEFDMLGP-----GSLIIVTSQDKQVLYCQVNEIKYVGLKHHSLQLQFSRCARGKD 310
QY 317 LAEGLLDIQK-----SRCLRNLM-KTPL---FVYTCAIOMGSEFHSHTQ 359
Db 311 VPQNLLELSMKFPVDYANGPNLALSLCKGNLKGTPDKMSVVL-----ELKRLHS 361
QY 360 TTLF---HTFPLDLIQKNH-----KHKGVAASDFIRSLDHG---DLALEGVFHKF 406
Db 362 DKIVYKLSSTDALSVSEKELFLDIVTFFRGANDVWQSLAGCGFPRRGIEALVDKSF 421
QY 407 DPELDVSSVNEDEVLLTTLG-----LCKYTAQREFPKYKFKHKSFOEYTAGR-- 453
Db 422 VTVAENRQVNA-NLIVDGLKIINDQDEICMCYRFVDASNSQSLIHKEIRESEQYED 480
QY 454 -RLSSLSLSHEPEVTKNGYLOKMWISDITTYSS----- 489
Db 481 VKAINLDSNLP---FKGHIAFOHMYMLRYLT--TYSSINPTKBDLPLPGDPOELPELR 536
QY 490 LLRYTCSSVEATRAVWKHLAAVYQHCGLLGLSLAK-RPLMRQESLSQVKTTOEILKA 548

```

Db 537 LHMTC---YPLASPPQNG--FOYLVELNMCSSKLKLM-----GGTNLEVLKR 582
OY 549 ININSVEGCIHLHYOESTSKSALSQEFAPFOGKSLYINGNIPDYLDFEFLPNCASA 608
Db 583 ITLS---CSVOLLN-----VDELOYSPIE-- 604
OY 609 LDEIKLDFYGAMASWEKAEDTGGIMHEBAPETIYPSRAVSLFFNWKQOFRTLEVTLRD 668
Db 605 ---KIDL-----KGCLELQSPDPT----- 620
OY 669 FSKLNQODITYLKITSSATSLSLQIKRCAGVAGSLSLV--LSICKNTYSILMEASPLTIE 727
Db 621 -----GOLQHLRIYDLSTCKTKKSF--PKVPISIR 648
OY 728 DEHHI---TSVTNKLTLSDHDLQNRPLGGLTDSLGNLKNLTKLIMDKKNEDEAIKLA 785
Db 649 -KHLAQGTGRDLSLNL-HSESOQLTKRL-----ENVSSNDHKKQVL 691
OY 786 GLKNLKKMCLFHLTHLSDIGEGMDYIVKSLS---SEPCDLEIQ-----LVSCCLSA 836
Db 692 KLDSS---HLGSLPDI-----VIFESLEVLDGSGSELEDIOGFPQYKRLYLAKTA 741
OY 837 VKTLAQL--HNLVKLSTLDSENVLEKDGNEALHELLDRNNVLEQLTALMLPGCDVQGS 895
Db 742 IKVPSLSLCHHISKLVKLDN-----ENCERLRLPMGKSNMXYLAVALKLS--GC----- 788
OY 896 LSSLKHLKEVLPOLVYKLGKLNMRITDTEIRILGAFEGKNLKNFOOLNAGNVSDDGWL 955
Db 789 --SNLENIKLP-----RNLKEIYLXAGTAVKE----- 813
OY 956 AFNG--VFENIKQLVFEDFSTKEFLPDPAVYKLSQVLSKTLPLQEARLVG 1004
Db 814 -FPSTLETLSEYVLD-----LENCKKIQGLPTGMSKLEFLVMKLKLSG 856

```

RESULT 14

T30553
 C:Species: Lycopersicon esculentum (tomato)
 C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 11-May-2000
 C:Accession: T30553
 R: Dixon, M.S.; Hatzixanthis, K.; Jones, D.A.; Harrison, K.; Jones, J.D.G.
 A:Title: The tomato Cf-5 disease resistance gene and six homologues show pronounced allelic variation
 A:Reference number: 220855; MUID:99030197
 A:Accession: T30553
 Status: preliminary; translated from GB/EMBL/DBJ
 Molecule type: DNA
 Residues: 1-1016 <DI>
 A:Cross-references: EMBL:AF053998; NID:93894392; PID:93894393; PIDN:AAC78596.1
 A:Genetics:
 A:Note: Hcr2-5D

```

Query Match 3.1%; Score 164; DB 2; Length 1016;
Best local Similarity 22.1%; Pred. No. 0.08;
Matches 147; Conservative 115; Mismatches 222; Indels 182; Gaps 38;

OY 459 LTHHEPEEYKKGYSLOKMWIS-DITSYSSILRTCGSSVATRVVMHIAAYOHGC 517
Db 203 LSGFTPEET---GYLRSILTKLSLDLN-----FLSGSIPASLGNNLSFLYLYNN 249
OY 518 LIGLSIAK---RPLWR-----QESLOSVKNTTEOEI----- 545
Db 250 QLSGSIPEELGYLSLTKLSLGINFLSGSIPASLGNNLSRDLVYNNKLSGSIPEIGY 309
OY 546 LKAININSVEGCIHLHYOES--TSKSALSOEFAPFOGKSLYINGNIPD--YL----- 595
Db 310 LRSLL---TYLDLGENALNGSIPASLGNNLSFLYLYNNQL---SGSIPREIGYLSRLTY 363
OY 596 FDFEFLPND---CASALDFIKLDFYGAMASWEKAADDTGGIMH-----EAPRET 642

```

```

Db 364 LDIGENALNGSIPASLGNNLSRDLVYNNKLSG--SIPREIGYLSRLTYLDIGENALNG 421
OY 643 YIPRAVSLFFENMKQOFRTLEVLRDFPSKLNKODITYLKITSSATSLSLQIKRCAGVAG 702
Db 422 SIPASLGNNL-----NNLFMYLYNNQLSGSIPEIGYL-----SLSLEIYLGNN---SLNG 469
OY 703 SLSVLSTCKNYSILMEASPLT---IEDERHITSVTNKLTLSDHDLQNRPLGGLTDSL 759
Db 470 SIPASLGNNLNNLEMYLYNNQLSGSIPEIGYLSLSTEL-----PLGNNLSNGSIPASL 523
OY 760 GNLKNTLTKLIMDKKNEDEAIKLAEG-LKNLKKMCLFHLTHLSD---GEGMDYIVKSL 815
Db 524 GNLNLSRLYLYNNQLS--GSIPASFGNNKRLDTL-----FLSDMDLIGELPSFCNLT 575
OY 816 SSE-----PC--DLEIQVSCCLSAANAVK-TLAQNLHNLVKLSITLDSN 858
Db 576 SLEVLVMSRNNLKGVPQCGNISDLHLS--MSSNSPREFLPSISNLSLTKLIDFGNR 633
OY 859 YLEKD-----GNEALHELLDRANVLEQLTALMLPGCDVQGSLSLKH---LEEVP-- 907
Db 634 NLEGAIPQEFNGNISLQVPMQN--NKLSG-TLPTNFSIGCSLISNHLGNELADEIPRS 690
OY 908 ---OLVYKLGKLNMRITDTEIRILGAFEGK--NPLK-----NEQOL 942
Db 691 LDNCKKLYLDLSDNQNDLTFPWLGLTEPLRYLRITSNKHGPIRSSGAEIMFPDLRII 750
OY 943 NLAGNRYSSDGMVLAFGVFNELKQLVFFSTKE-----FLPDPAVY-----RKLQVYL 991
Db 751 DLSRNAFSQD---LPSLSFELHAKMRTVDKTMEEPSYESYDSDSVVYVYKGLEIVRLI 807
OY 992 SKLTFPL 997
Db 808 SLVYII 813

```

RESULT 15

A40264
 C:Species: Xenopus laevis (African clawed frog)
 C:Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 19-Jan-2001
 C:Accession: A40264; S18764
 R:le Guellec, R.; Paris, J.; Couturier, A.; Roghi, C.; Philippe, M.
 A:Title: Cloning by differential screening of a Xenopus cDNA that encodes a kinesin-related protein Eg5
 A:Reference number: A40264; MUID:91246212
 A:Accession: A40264
 A:Molecule type: mRNA
 A:Residues: 1-1060 <LE>
 A:Cross-references: EMBL:X54002; NID:g64869; PIDN:CAA37950.1; PID:g64870
 A:Note: The authors translated the codon GCT for residue 784 as Leu
 C:Superfamily: kinesin-related protein Eg5; kinesin motor domain homology
 C:Keywords: ATP; coiled coil; microtubule binding; mitosis; nucleotide binding; P-loop
 F:12-358/Domain: kinesin motor domain homology <KMO7>
 F:98-105/Region: nucleotide-binding motif A (P-loop)
 F:104/Binding site: ATP (Lys) #status predicted

```

Query Match 3.1%; Score 164; DB 1; Length 1060;
Best local Similarity 20.3%; Pred. No. 0.085;
Matches 214; Conservative 155; Mismatches 383; Indels 302; Gaps 52;

OY 24 TDDLFWYNNVLRREVNICEKVEDAARGIT---HMLIKGSESCNLF--LKSLEK 76
Db 114 SDEFTW-----EQPLAGITPRTLQHFKEKLSNGTFESVYVLSLEYI 157
OY 77 NYPLFQDLN-----GQSL-----FHQTSQEG---DLDLADQLKD-LYN-----T 111
Db 158 NEELFLLSPDPVGRGLQFDDPRKKRGVYIKGLEISVHNKDEYVNLERGAARRKTA 217
OY 112 PSFLNYPRLGED--IDILFNLKSTFEPRVLYMKRQDHHNRVEQLTGLGALQALOSPCTIEG 169
Db 218 STLMAVSSSHSVSYTTHMKETVYG-----EELVYKICKLNLVD---LAG 261

```

QY 170 ESGKSKSTLLORIAMLMGSKCALTKRFKVFPL-----RLSRAQGLFETLCDDLDLP 224
DB 262 SENIGRSGAVDKRAREAGINOSLLTLCRVITALVERTPHIPIRESKLTIRLQDSL----- 317
QY 225 GYIRKQTFMAML-----LKLQKRVLFLLDGYNEFKP-QNCPEI-----EALIKENHRRKN 273
DB 318 GGRRTSIIATVSPASINLEETV-STLDYANRAKSIMNKPEVNQKLTKKALIKE----- 370
QY 274 MYIVTTTECRLHITROFAGLTAEVG-----DMTEDSAQALIREVLKELAGLLQIQK 327
DB 371 -----YTEIEIERLKRELAAREKNGVYLSENEYEOLOGKVLQSEPMITEYTEKITAMEEE 425
QY 328 SRCLENNLTKTLPFVVYTCA--IOMGESEFHSHTQTLTFHTFYDLLOKNNKHKGAASD 385
DB 426 LKSISELPAADKKKELECTTIIQCKEKEL-BETQNH-----QESKEQ---LAQES 472
QY 386 FIRSIDHCGDLALBEGVSHKDFEFLQDYSSVNEVDVLLTGLCKTYAORFKPKYKF-FHK 444
DB 473 FVVSAPFETTERKKLHGT-ANKL-----LSTVARETRDVSGLHEKIDRRKAVDQHNFOVHE 525
QY 445 SFOEYTAGRRISLLTSHPEPVTKNGVLOKMV-SISDITSTVSSLRYTCGSSVEATR 503
DB 526 NFAE-QMDRRFSVIORTVDYSV-KQGMLDFTYNSIDDLGASSSRSLATAS----- 576
QY 504 AVMKHLAAVYOHGCLLGIKAPLMROESLOQSVKNTTEOEILKAININSFEVCGIHLYQ 563
DB 577 AVAKSFASV-----OETVKOVSHVEELKQETLS----- 607
QY 564 ESTKSALSQEFKFAFQCKSLYINSGNIP--DYLFDFEHLPCNASALDFI--KLDFYGG 619
DB 608 -SOAKGDLQOOLAAHARTGLEALRSDLPVYAVAVLDLNSHLSHCLQNFLLIYADKIDSHE 666
QY 620 AMASWEKAEPDTGGIHMEAPETIIPSRASVLEFNMKOEFTLE----- 663
DB 667 DMNSF--FTEHSRSLHKLRLDSSSALSISQSEYLSKEDIATQAQSMHSEGVNLLISSLQN 724
QY 664 -----VTLRDFS-----KLNK-----ODITYLG-----KIFSSATSLR 691
DB 725 QLNILGMEFQOQFSGFLSKGGLQKSVGLQDDDLVSSSAIECISSHKKLAEGSQDYA 784
QY 692 LOIKKACAVAGSLVSLSTCKNIYSLMEVASFPLTEDE-----RHITSVTNLKLSIH 744
DB 785 VEIROLAG--SNMSTLEESSKQCEKLTSSINTISOESQWCESAGOKIDSYLEQVCYLH 842
QY 745 DLQN--ORLPGGLTDSLGN-----LKNLTKLIMDNIKMNEDATKL 783
DB 843 SSKHLQNLHKGVEDSCSSVVEITDRVNAQRAEKAELTSLVEQVRDQEMVGEQRLLE 902
QY 784 AEGLNK-LKKMCLFHLTHL-SDIGEG-----MDYIVKSL--SSEPCDL-----EE 824
DB 903 QEOVQSGLNKHYSLKELRNDVPTGCTTPQRDVAIPSLVKTFRDVLLEQFRQOQOEX 962
QY 825 IQLVSCCLSANAVK-----IIAQNLHNLVKLSILDSENYLEKDG----- 864
DB 963 LESISSVIS-EAVEPPVPODSLDEDPVAVNDVSYSERSCIDLSMTQCKGKIRFFQOKK 1021
QY 865 -----NEALHELIDRNNVLEQ-----LTALMLP 887
DB 1022 ALRKEKENRGNTTLERSKIMDEVQALTKSKLP 1055

Search completed: March 25, 2002, 10:56:44
Job time: 104 sec

THIS PAGE BLANK (USPTO)


```
1743 .....CAMGACCTGTGCTGCTGCTGAGCCTCTTGGCACTTGA 1782
162 InseProCysIleIleGluGlyIleuSerGlyGlySerThrIleu 178
1783 ACCTGTCATGTCGTGAGGAGTGAAGTGAAGTGAAGAGCGCTC 1832
179 LeuGlnArgIleAlaMetLeuTrpGlySerGlyLysCysAlaLeuTh 195
1833 CTGAAGAAATAGCTTTCTGTGGGCACTGGAGCTGCTCCCTGTTAAA 1882
195 rLysPheLysPheValPhePheLeuArgLeuSer.....ArgAlaIng 210
1883 CAGGTCACAGCTGTTTCTACTCTCTCCCTAGTCCACGACGACAGC 1932
210 LysIleuPheGluThrLeuCysAspIleuLeuAspIleProGlyThr 226
1933 AGGGGCTGGCAGATCATCTGTGACGAGCTCTAGAGAAAGAAAGATCT 1982
227 IleArgLysGlnThrPheMetAlaMetLeuLeuLysLeuArgIleArg 243
1983 GTTACTCAATGTGATGAGGACATTAATCCAGAGTTAAAGAAATCAGGT 2032
243 llePheLeuLeuAspGlyThrArgIleuPhe.....LysProGlnAsn 258
2033 CTTATTCCTTTTATGATGACTACACAGAAATATGTTCATCCCTCAA... 2078
258 yspProGluIleGluAlaLeuIleLysGluAsnIleArgPheLysAsnMet 274
2079 .....GTCAATGAGAAACTGATTCAAAAAACCACTTATCCCGACCTGC 2123
275 ValIleValIleThrThrThrGluCysLeuArgHisIleArgGlnPheG 291
2124 CTATTGATGTCGTGCTGACAAACAGGCGCCAGGACATCCGCGATACCT 2173
291 yAlaLeuThrAlaGluValGlyAspMetThrGluAspSerAlaGlnAla 308
2174 AGAGACCATTCATAGATCCAGACATTTCCCTTTATATATCTGTCTGA 2223
308 euIleArgGluValIleuIleLysGluLeuAla.....GluGlyLeuLeu 322
2224 TATTACGGAACCTCTTTTACATATATGATGACTCTGCGAAAGTTATG 2273
323 LeuGlnIleGlnLysSerArgCysLeuArgAsnIleuMetLysThrPro 339
2274 GTTTACTTTGGAAAGAACCAAGTTTGGCAAGATACGAAAGTCCCTCT 2323
339 uPheValIleThrCysAlaIleGlnMetGlyLysSerIleuPheHis 356
2324 CTTTGTGGCGGCGATCTGTCTCATTTGTTCAATCCCTTTGACCAT 2373
356 erHisThrGlnThrThrLeuPheHisThrPheTyArgPheLeuIleGln 372
2374 CTTTGTGATGATGCGCTGTTTCAAGTCATATGAAACGCTTCTCTTA 2423
373 LysAsnLysHisLysHisLysGlyValAlaAlaSerAspPheIleArg. 388
2424 AGGAACAAA.....GCGACACTGAAATTCACAAAGC 2455
389 .SerLeuAspHisCysGlyAspLeuAlaLeuGluGlyValPheSerHisL 405
2456 AACGTGTCTCCTGCTGTGCTGAGTGGCTTGAAGGGTTTTCATGTT 2505
405 yspPheAspPheGluLeuGlnAspVal.....SerSerValAsnGluAsp 419
2506 GCTTGTGATTAATGATGATGATCTCGAGAAAGAGGGGTGATGAGAT 2555
420 ValLeuLeuThrThrGlyLeuLeuCysLysTyThrAlaGlnArgPheL 436
2556 GAAGATTAACCATGCTGATGAGCAAAATTTACAGCCGACAGACTAAG 2605
436 sProLysTyLysPhePheHisLysSerPheGlnGlyThrAlaGlyA 453
2606 ACCATTTCACGGGTTTAACTGCTGCTCCCAAGAAATTTCTTGGGGGA 2655
453 rGArgLeuSerSerLeuLeuThrSerHisGluProGluGluValThrLys 469
2656 TGAGCTCTGATTAAGACTCCTGATTCAGATAGCAGAAACATCAAGATTG 2705
470 GlyAsnGlyTyLeuGlnLysMetValSerIleSerAspIleThrSerTh 486
2706 GGACTGTATCAATTTGAAACAAATCAACCTCACCATGATGACTGTAGGCG 2755
486 rTySerSerLeuLeuArgTyThrCysGlySerSerValGluAlaThrA 503
2756 CTACACATTTTGTGAACATGTC.....TCCAGCTCCTTCACAA 2799
503 rGAla.....ValMetLysHisLeuAlaAlaValTyrgLHisGly 516
2800 AAGCAGGCGCCAAATGTGTCTCATTTGTCATTTAGTGTGAATAC... 2846
517 CysLeuLeuGlyLeuSerIleAlaLysArgProLeuThrArgGlnIle 533
2847 .....AAAGATC 2854
533 rLeuGlnSerValLysAsn..... 539
2855 ATTGAGCATATATCTGAAATATGATGACTACTTAAGCAGCCAGAGAA 2904
540 .ThrThrGluGlnGluIleLeuLysAlaIle.....AsnIle 551
2905 TTTCACATCAGATGACAGTACTTACGAGGATGTGCAAAATTTGTCCAA 2954
552 AsnSerPheValGluCysGlyIleHisLeu..... 561
2955 GCTTACTTTCAATGTTTCAGAACATTTACTGTTCTTGCCCTGAAAC 3004
562 ....TyrgGlnIleuSerThrSerLysSerAlaLeuSerGlnIlePheG 577
3005 TGCTTATCAAGCAACT...GTTGCTGCTGTTCCTTCATTTGTTTGC 3051
577 LapePheGlnGlyLysSerLeuTyrrIleAsnSerGlyAsnIleProAsp 593
3052 AATTCCTTCAAGGAGAAACACTGACTTGGTGGCTTAACCTTA...CAG 3098
594 TyrLeuPheAsp..... 597
3099 TACTTTTTCACCAACCAAGAACTGTTCATTTGTGAGCAGACATTCAC 3148
598 .....PhePheGlnHisLeuP. 603
3149 CTCAATACGAGGAATATAGACATCACCCAGACACATTTTCAGTCTGG 3198
603 roAsnCys.....AlaSerAlaLeuAspPheIleLysLeuAspPheTy 617
3199 AAACATGTTTGAACAATACACAGTGGCCACACTATAGATCAGAGACTATG 3248
618 GlyGly.....AlaMetAlaSerTyrGlu..... 625
3249 TCTGCTTTTGAACCTATGATGATGATGAGGAGAAATTTAGCTGAAAAA 3298
626 .....LysAlaAlaGluAsp 631
3299 GGATATATTAAGAGCTATATGATATCCAGGACGACATTCACACAGCC 3348
631 hrGlyGlyIleHisMetGluGluAlaProGluThrTy...IleProSer 646
3349 TTAGTACTGCTATTTGAAACTTCTCCAAAGCAGTACAAAGATTCCTCC... 3395
647 ArgAlaValSerLeuPhePheAsnTyrLysGlnIleuPheArgThrLeu 663
3396 .....TGTCTGA 3403
663 uValThrLeuArgAspPheSerLysLeuAsnLysGluAspIleThrTyL 680
3404 AGTCGATGTGAATGATGATGATGTTGTGAGCCAGAGATGTGCTGAGATTC 3453
```



```

680 eucllystlepheserlatlnserleuabglnlleylsarg 696
    || :|||||:.....:.....:||||:||||:
3454 TAATGCAGTTTCTCAGCTTCACAGCCGATCGAATCATTTAAACAC 3503
697 cysalaglvalaglyserleuSerleuValleuSerThrcysLys.. 712
    ||| :|||: :|||: :|||: |||
3504 AGCAGAGGCTTTATAGAAAAGCATCCGCCACAGCTTGAGAGCTGTAAAGC 3553
713 .AsnIlyrSerLeuMetValgluaIsarProLeuthrllegluaSpG 729
    :||: :||: :||: :||: :||:
3554 CTCTGCACACAAGTCTCCATAGCAAGTTGGAACTCACAGCCAGCCGAAC 3603
729 luarghlslrherServalthrAsnleuLyThleuSerlleHis... 744
    :||: :||: :||: :||: :||:
3604 AGGAAGCTGTCTTCACCCCTGCCTCCGTGAATCTCTTGAAGTCCAGAGG 3653
745 .....AspleuglnasnglnArgleuProglyLyLeuThrAspSerle 759
    :||: :||: :||: :||: :||:
3654 ACAATTCACATCACAAGACCMAATCTTTCTTAAT....CTGGATAAGTT 3697
759 uGlYasnleuLySasnleuThrLysleuIllemetaspasnilleysmet. 775
    : |||||:||||: |||||: |||||:
3698 CCTGGCCCTGAAAGAAGACTGTCTGTGAGCTGGAGGCAATATAATATT 3747
775 .....:.....:.....:.....:.....:..... 775
3748 TTTCAGTCATTCTCGAGAAATTTCCAACTTCACCAATATGAGAGAAATTA 3797
776 .....:.....:.....:.....:.....:..... 3847
3798 TTGATCCAAATTTCAAGCTGAGTAGATGATCTTCCAACATGAGTGGCAGTTT 3847
787 ulYasnleuLySlYsmetCysleuPheHISleuthrHISleu..... 801
    | :|||: :|||: :|||: :|||:
3848 GGCAATATTTATTCTCGAAGATATTAATCTTGAAGCCAGCAGCATTTTC 3897
802 .....SerapllleglygluylmetaspyrilevallysSerleuSer 816
    :||: :||: :||: :||: :||:
3898 CTGATGAGGAAACATCAGAAAAATTTGCCCTCATATTTAGCTTCTCTTGT 3947
817 SerGIUProCyaspLeugluGluIleGlnleuValserCysyleuse 833
    :|||: :|||: :|||: :|||:
3948 .....AACCTGGAAGAAATGATTCCTCTTACTGGGAGATGAT 3985
833 rAlaSnalavalYslleuAlaglnasnuIsanleuValysl 850
    :||: :||: :||: :||: :||:
3986 TTAATGAGAGGCCAAACATGATCATCCACAGAGTGCAGCAGCTTCAATTGC 4035
850 euSerlleu 853
    || :|||
4036 TCCGAGTCTC 4046

seq_name: /cgn2_6/pltodata/2/ina/dh_COMB.seq-US-08-930-996A-1

seq_documentation_block:
? Sequence 1, Application US/08930996A
? Patent No. 6100449
? GENERAL INFORMATION:
? APPLICANT: FLUHR, Robert
? APPLICANT: ESHED, Yuval
? APPLICANT: ORI, Naomi
? APPLICANT: PARAN, Ilan
? APPLICANT: ZAMIR, Daniel
? TITLE OF INVENTION: A GENE FAMILY FROM THE I2 FUSARIUM RESISTANCE
? TITLE OF INVENTION: LOCUS OF TOMATO AND USE THEREOF FOR TRANSFORMATION AND
? NUMBER OF SEQUENCES: 12
? CORRESPONDENCE ADDRESSES:
? ADDRESSEE: BROMIDY AND NEIMARK
? STREET: 419 Seventh Street, N.W., Suite 300
? CITY: Washington
? STATE: D.C.
? COUNTRY: USA
```

[illegible]

136 GluPro..... 137
 749 TCACCAAAACAGAACTAGACACCTTCAACTTCTTGGTTGATGATTC 798
 138 ...ValLeuTPArgLysAspGlnHisHisArgValGluGlnLeu 153
 799 TGGTATCTTTGGAAGAAAGTAATAGAAATTTGGTGGCGTTGT 848
 153 hrLeuAsnGlyLeuGlnAlaLeuGlnSerProCysIleIleGluGly 169
 849 TGCTATGATGATACAAAGCAAAAAATGGCTGTAGTTCCTATTGTGGCA 898
 170 GluSerGlyLysGlyLysSerThr..LeuLeuGlnArgIleAlaMetLeu 186
 899 ATGGCGCGCATGGGTAGACACACTGTCTAAAGCCGTTACAAATGATG. 947
 186 rpeLysGlyLysCysLysAlaLeu.....ThrLysPheLys 198
 948AGAGAGTCGAGAAACATTTGGTTGCACAGCTGTGGTTGTG 989
 199 PheValPhePheLeuArgLeuSerArgAlaGluGlyLeuPheGluTh 215
 990 TTTCTGAGCATATGATGCTTTTCAGAAATACCAAAAGCTTACTTCAAGAA 1039
 215 rLeuCysAspGln.....LeuLeuAspIleP 224
 1040 ATTTG...GATCACTGACTGAGGCTGATGACAACTTATCATCAGTACA 1086
 224 rGlyThrIleArgLysGlnThrPheMet..AlaMetLeuLysLeu 240
 1087 AGTCAAATTAAGGCTGATGACATCTTAATCAAGTCAAGTCAAAATGGA 1136
 240 rGln.....ArgValLeuPheLeuAspGly...Tyr 250
 1137 AGGAAAGCTGAATGAAAAAGTTTCTGTGTGCTTGAATGAGCTGTGG 1186
 251 AsnGluPheLysProGlnAsnCysProGluIleGluAlaLeuIleLysG 267
 1187 AATGAT.....AATTTCTCTGAGTGGGATGACTTG..... 1216
 267 uAsnHisArgPheLysAsnMet.....Y 275
 1217AGAAATCTTTTTCACAAAGGGAATAGAGAACTAGA 1253
 275 allLeuValThrThrThrGluCysLeuArgHisIleArgGlnPheGly 291
 1254 TCATTGTAAAGACACGTAAGAGAGTGTGCTTGATGATGATGATGGG 1303
 292 AlaLeu..... 293
 1304 GCAATCTACATGGGAATTTCTGTCTAGTGAAGACCTTGGGCTATTCAA 1353
 294ThrIleG 296
 1354 ACGACATTCATTAGAGCACAAGATCCCAAGAAATCCAGAAATTTGAAG 1403
 296 lValAlaGlyAspMetThrGluAspSerAlaGlnAlaLeuIleArgGlyVal 312
 1404 AGGTTGAAAAACAATTCAGACAGAGTGCAGAAAGGTTG.....CCTTTA 1447
 313 LeuIleLysGluLeuAlaGluGlyLeuLeuGlnIleGlnLysSerAr 329
 1448 GCCTTAAACACACTTGCT...GGTATGTACGCGCAAAATCAGAGGTGGA 1494
 329 gCysLeuArgAsnLeuMetLysThrProLeuPheValIleThrLys. 345
 1495 TGAGTGGAGAAACATTTTACGAGGTGAATGTGGAGCTTCCAAAGTTGT 1544
 346AlaIleGlnMetGlyGlnSerGluPheHisSer 356
 1545 CGAATGGTATATACACGCGCTAATGTTGACGTCATGATCTCCCTGCA 1594

357 HisThrGlnThrThrLeu..... 362
 1595 CATTTAAAGCAATGTTGGCTTATTGTCAATATATCCAAAGATATACA 1644
 362 362
 1645 ATTTGCCAAAGACAGCACTTATTCACCTGTGGATTGCTAATGCTGTGTAC 1694
 363PheHis.....Thr 365
 1695 ATCAGTTTCATTCGGGTACCAACATCTTATCGAGTTGAGATCAAGATCA 1744
 366 PheTyrAspLeuLeuIleGlnLysAsnLysHisLysHisGlyValAl 382
 1745 TTGTTCCAAATAGCCCTCAGACCTTCTGAAAGAGACGTAGAGAAATCTT 1794
 382 aAlaSerAspPheIleArgSerLeu.....AspHisC 393
 1795 AATGCATGACCTTGTCATATGATTTGGCAAAATTCATCTTCAATCATTT 1844
 393 ys.....GlyAspLeuAlaLeuGluGlyVal... 401
 1845 GTATAAGGTTGGAAGATATACAAAGATCGCATATGTGGAACATGTGCG 1894
 402PheSerHisLysPheAspPheGluLeuGlnAspValSer 415
 1895 CACATGTCCTATTCATCAATAGACAAAGATGCTGAGTTGAGAAATGAAATC 1944
 415 rValAsnGluAsp.....ValLeuLeuThrThrGlyLeuLeu 428
 1945 ACTTTTAAATCAGACGAGCTGAGGACATTTACTTCCAAATGCATTCAGT 1994
 428 yLysTyrThrAlaGlnArgPheLysProLysTyrLysPhePheHisLys 444
 1995 TCCATTACTCAAAAAAACTAAGCAAG.....AGGTTGTCATTAAC 2035
 445 SerPheGlnGluTyrThrAlaGlyArgArgLeuSerSerLeuThrSe 461
 2036 ATACTGCTACACTA.....AGATCTGTGAGGCACTATCATTTGTC 2076
 461 rHisGluProGluGluValThrLysGlyAsnGlyTyrLeuGln...LysM 477
 2077 TCATTACCAAGATTGAGGTGTTGCCAAATGACTGTTTATCAATTAAGC 2126
 477 etValSerIleSerAspIleThrSerThr..... 486
 2127 TCCTCAGATTTTGGACCTTCTGACACATGCAATGATTACAAAGTCCGGAT 2176
 487TyrSerSerLeuLeuArgTyrThrCysG 496
 2177 TCCATTTTGTGTGTATTACTTGAAGACACTTCTCCTGCATCTGTGTA 2226
 496 ySerSerValGluValArgAlaValMetLys..... 507
 2227 ATATCTTGAGAGCTACCGCTGCAAGATGAGAGAGTTGATTACTTGGCTC 2276
 508HisLeuAla 510
 2277 ATCTTGACATAGCAACACATCGGCGCTTGAAGATCCGACATCACTAGC 2326
 511 AlaValIyrGlnHisGlyCysLeuLeuGlyLeuSerIleAlaLysArgPr 527
 2327 AGGTTGAAAGCCCTCAAGTGTGTGTGGAGCCCAAGTTCTTGTAGGTGG 2376
 527 oLeuTyrArgGlnGluSerLeuGlnSerValLysAsnThr...ThrGlu 543
 2377 T...TGGAGATGGAATATTTGGGTGAAGCACCACATCTTATATGATCTC 2423
 543 lngluIleLeuValAlaIleAsnIleAsnSerPheValGluCysGlyIle 559
 2424 TATCAATTTCTAGACTTGAAAAATGTGGTGTATAGAGGAAGCTGTGAAG 2473
 560 HisLeuTyrGlnGluSerThrSerLysSerAlaLeuSerGlnGluPheG 576

```
2474 GCAAGATGAGGAGAGATCATTTGAGCAATATTCATTTGAGTGGAG 2523
576 uAlaPhepHeNgly..... 581
2524 TGAACCATTTAGTGTGACAAATTCACAAACAGAACATACCTTGATG 2573
582 .....LysSerLeuTyrIle.....Asn 587
2574 AGCTACGCCACATAAAACATTAAAGCAGTGAATTCACGTGATATACA 2623
588 SerGlyAsnIleProAspTyrLeuPheAsp..... 597
2624 GGGACAACCTTCCAAACTGGAGCTGATCCTTTGTTTAAGCTGTG 2673
598 .....PhepHeGluHisLeuProAsnGlyAlaSer..... 607
2674 GCATTTGTATCTTAGAACCTGCAGAGACTGTACTCTTGCACGACACTAG 2723
608 .....AlaLeuAspPheIleLysLeu..... 614
2724 GACAACCTCCTGTTGGAATTCCTTCATTAGAGGATGCATGGGATA 2773
615 .....AspPheTyrGlyGlyAlaMetAlaSerTrpGluLys 626
2774 AGAGTGTGACAGAGAGTTCATGAC.....AGATTGCTCTCCAAAAA 2817
626 sAlaAlaGluAspThrGlyGlyIleHisMetGluGluAlaProGluThr 643
2818 GCCTTTTAACCTCTTGCAACCTAGATTGAAGATATGCTCGAA..... 2863
643 yIleProSerArgAlaValSerLeuPheAsnTrpLysGln..... 657
2864 .....TGGAAAGCAATGCGAC 2878
658 .....GluPheArgThrLeuGluValThrLeuArgAspH 669
2879 ACATAGTAATTTGAGAGTTCCTACACTTGAG..... 2911
669 eSerLysLeuAsnLysGlnAspIleThrTyrLeuGlyLysIlePheSer 686
2911 ..... 2911
686 eAlaThrSerLeuArgLeuGlnIleLysArgCysAlaGlyValAlaGly 702
2912 .....AACTTTCATTTAAATTTGCCCTGAGCTCAGTTGG 2947
703 SerLeuSerLeuValLeuSerThr.....Cys 711
2948 GAGATACCCATCCAAATTTCAAGTTTAAAGTTAGATATATGTCATG 2997
711 sLysAsnIleTyrSerLeuMetValGluAlaSerProLeuThrIleGlu 727
2998 TAAGTCTGTACTCTCTCTTTTCTTAGCATCTACCTACCTGGAAGA 3047
728 .....Asp 728
3048 GATAAAGATATCTGTTGCCCAAAATTTGAATTTGAGGCGCAGTTGCT 3097
729 GluArgHisIleThrSerValThrAsnLeuLysThrLeuSerIleHis 745
3098 GAGATGTTTGTGAGTATTTGAGTGTGATTTGATGTTGTTGATGATGA 3147
745 pleuGlnAsnGlnArgLeuProGlyGlyLeuThrAspSerLeuGlyAsnL 762
3148 TATATCACCTGTAGTTTCTCCACACGCAAGTCAATTTGATTTGGAAT 3197
762 eLysAsnLeuThrLysLeuIleMetAspAsnIleLysMetAsnGluGln 778
3198 GGCACAACGTTTCTAGGTTTTCATCTCAGTCCACATCAAGTCTCCAT 3247
779 AspAlaIleLysLeuAlaGluGlyLeuLysAsnLeuLysLysMetCysL 795
```

```
3248 .....ATTCGGAATTTGAAAAACTCTGCAT 3273
795 u.....PheHisLeuThrHisLeuSerAspIleGlyGlu 807
3274 GGCATGTGAGAGACGCGCCAGCTGACGTCATGAAATATTGGGGA.... 3319
807 LysMetAspTyrIleValLysSerLeuSerGluPro.....Cys 820
3320 .....TGTAAAGACCTCAAGTCTCTCCAGACCTCTTCCA 3355
821 AspLeuGluGluIleGlnLeuValSerCysLysLeuSerAlaAsnAlaVal 837
3356 TCTCTCAAGAACCTGCGACTGACTTATGT.....CC 3387
837 LysIleLeuAlaGlnAsnLeuHisAsnLeuValLysLeuSerIleLeu 854
3388 AGAAATAGAGAGGAATTTGCCCTTCAT.....TTACAAATTCCTG 3428
854 sPheSerGluAsnTyrLeuGluLys.....AspGlyAsnGluAlaLeu 868
3429 ATATC.....AGATTTCGAAGAACTGTGATGCGCGAAAGAGTGG 3472
869 HisGluLeuIleAspArgMetAsnValLeuGluGlnLeuThrAlaLeu 885
3473 CAT.....TTACAGAGACTCACAGAGTTATG 3498
885 LeuProTrp...GlyCysAspValGlnLysSerLeuSerLeuLeu 901
3499 GATCAAAACATGATGGAGTGAC.....G 3521
901 yHisLeuGlu.....GluValProGlnLeuValLysLeuGlyLeuLys 915
3522 AACATTTGACATTTGGAGTTCCTCTCTATTCAG..... 3559
916 AsnTrpArgLeuThrAspThrGluIleArgIleLeuGlyAlaPheMetG 932
3560 .....AGACTTATTCATTCATCTGAACATCTAAGACGCAACATCT 3603
932 yLysAsnProLeuLysAsnPheGlnGlnLeuAsnLeuAlaGlyAsn.... 947
3604 CAAGAC...CTACCTCTCTCAATTTCTACGTATTTGTTGAATTTAT 3650
948 ..ArgValSerSerAspGlyTrpLeuAlaPheMetGlyValPheGluAsn 963
3651 CTCAGTTTACGTCACAGGCCAACTTCTCTCTTCTCACCCTCAGTTG 3700
964 LeuLysGlnLeuValPhePheAspPheSerThrLysGluPheLeuProAs 980
3701 CTTCAAACTCTCAAAATCTGAATTTCTTAATCTTCATACCTACCTGA 3750
980 proAlaLeuValArgLysLeuSerGln...ValLeuSerLysLeuThr 996
3751 ATCAGCACTGCCCTCCCTCTCTCACCCTGATCATCTCAATTTGCCCTA 3800
996 heLeuGlnGluAlaArgLeuValGlyTrpGlnPheAspAspAspLeu 1012
3801 ATCTCCAACTCCTTCATTTAAAGGATGCGCTTCCCTCTCTACACTA 3850
1013 SerVal 1014
3851 TCAATT 3856
```

seq_name: /cgn2_6/ptdata/2/ina/6B_COMB.seq.us-09-180-439-1

seq_documentation_block:

- Sequence 1, Application US/09180439
- Patent No. 622532
- GENERAL INFORMATION:
- APPLICANT: Dixon, Mark S
- APPLICANT: Hatixanthis, Kostas
- APPLICANT: Jones, David A
- APPLICANT: Jones, Jonathan DG
- TITLE OR INVENTION: Plant pathogen resistance genes and uses thereof

```
FILE REFERENCE: 620 - 53
CURRENT APPLICATION NUMBER: US/09/180,439
CURRENT FILING DATE: 1998-12-06
EARLIER APPLICATION NUMBER: PCT/GB97/01249
EARLIER FILING DATE: 1997-05-08
EARLIER APPLICATION NUMBER: GB 9609681.3
EARLIER FILING DATE: 1996-05-09
EARLIER APPLICATION NUMBER: GB 9619924.5
EARLIER FILING DATE: 1996-09-24
NUMBER OF SEQ ID NOS: 10
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 1
LENGTH: 3979
TYPE: DNA
ORGANISM: Lycopersicon esculentum
US-09-180-439-1

alignment_scores:
Quality: 172.00      Length: 639
Ratio: 0.533         Gaps: 38
Percent Similarity: 50.548 Percent Identity: 23.005

alignment_block:
US-09-697-089-2 x US-09-180-439-1 ..

Align seg 1/1 to: US-09-180-439-1 from: 1 to: 3979

459 LeuHrSerHisGluProGluGluValThrLysGlyAsnGlyTyrLeuG1 475
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1260 CTTTCTGCGCTTATTCCTGAGAAATA.....GGTTACCTTAG 1297
475 nlyMetValSerLeuSer...AspIleThrSerThrTyrSerSerLeuL 491
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1298 GTCCTTACTAGACTACTTGGATATCAAC..... 1328
491 euatGlyThrCysGlySerSerValGluAlaThrArgAlaValMetLys 507
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1329 .....TTTCTTAGTGGTCCATCTCTGCTCATTTGGGGAATCTGAC 1370
508 HisLeuAlaAlaValTyrGlnHisGlyCysLeuLeuGlyLeuSerIleAl 524
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1371 AACTTGTCTTTTGTATCTTTACATATATGAGCTTTCTGCTCTATT. 1418
524 alyArgProLeuThrParGlnGlnSerLeuGlnSerValLysAsnThrT 541
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1419 .....CCTGAAGAATAGTTACTTACCTAGTGACACTTA 1449
541 hrGluGlnGluIleLeuLysAlaIleAsnIleAsnSerPheValGluCys 557
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1450 CTAG.....CTATCTTTG 1463
558 GlyIleHisLeuTyrGlnGluSerThrSerLysSerAla..... 570
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1464 GGTATCACTTCTTAGTGGTTCCTCTCTTCATTTGGGGAATCTTAA 1513
571 .....LeuSerGlnLupheGluAlaPheheGlnGlyLysSerLeuTyrI 586
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1514 CAACCTGTCTAG.....TTGATCTTTTCA 1539
586 leasn.....SerGlyAsnIleProAsp.....TyrLeu..... 595
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1540 ATAATAGCTTTCGCTCTATCTCGAATAATAGTTACCTACGAGCTCT 1589
596 .....PheaspPheheGluHis.....LeuProAs 604
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1590 CTTACTTACCTAGATTGGTGAGAAATGCTTTATGCTTATTCCTTC 1639
604 nCysAlaSerAlaLeuAspPheIle...LysLeuAspPheTyrGlyAla 620
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1640 TTGATTTGGGGAATCTAAACAACCTTGTCTAGTGTGATCTTACAAATA 1689
620 lametAlaSerTrrpGluLysAlaIleGluAspThrGlyIleHisMet 636
```

```
1690 AGCTTTCTGCG.....TCATTCCTGAGAAATAGTTACCTAGGCT 1733
637 .....GluGluAlaProGluThrTyrIleProse 646
1734 CTTACTTACCTAGATTGGTGAGAAATGCTTTATGCTTATTCCTGCG 1783
646 rArgAlaValSerLeuPhePheAsnTrrpLysGlnLupheArgThrLeuG 663
1784 TTCAATGGGGAATCTG.....ACAACCTGTTATGTTGT 1818
663 luValThrLeuArgAspPheSerLysLeuAsnLysGlnAspIleThrTyr 679
1819 ATCTTTACAAATATACGCTTTCGCTCTATTCCTGAGAAATAGTTAC 1868
680 LeuGlyLysIlePheSerSerAlaThrSerLeuArgGlnIleLysAr 696
1869 CTG.....AGTTCTTACTGAACTATATTTGGGTAATAC. 1904
696 gCysAlaGlyValAlaGlySerLeuSerLeuValLeuSerThrCysLysA 713
1905 .....TCCTTATGAGCTCTATTCCTGTTCAATGCGGAATCTGAA 1947
713 snIleTyrSerLeuMetValGluAlaSerProLeuThr.....Ile 726
1948 ACTGTTTATGTTGATCTTACATATATACGCTTTCGCTCTATTCCT 1997
727 GluAspGluArgHisIleThrSerValThrAsnLysThrLeuSerIle 743
1998 GAAGAATAGTTACTGAGAGTCTCTTACTGAACTA..... 2033
743 eHisAspLeuGlnAsnGlnArgLeuProGlyGlyLeuThrAspSerLeuG 760
2034 .....TTTGGGTAATAACCTCTTAATGCGCTATTCCTGCTCATTTG 2079
760 lyAsnLeuLysAsnLeuThrLysLeuIleMetAspAsnIleLysMetAsn 776
2080 GGAATCTAAACAACCTTGTAGGTGATCTTACAAATATACACCTTCT 2129
777 GluLysAspAlaIleLysLeuAlaGlyLys.....LeuLysAsnLeuLys 792
2130 .....GGCTATATCCGCTTCATTTGGCAATATGAAATCTGCAAC 2173
792 smetCysLeuPheHisLeuThrHisLeuSerAsp.....IleGlyG 806
2174 TCTG.....TTTCTCAGATGATACGATCTCATTTGGG 2205
806 luGlyMetAspTrrIleValLysSerLeuSerSerGlu..... 818
2206 AAATTCCTTCAATTGTGTGCAATTTGACATCAGTGAAGTGTGTATATG 2235
819 .....ProCys.....AspLeuG1 823
2256 TCGAGAAACAATTTGAGAGGAAAGTTCCGCAATGTTGGGTAATATACG 2305
823 uGluIleGlnLeuValSerCysLeuSerAlaAsnAlaValLys...I 839
2306 TGACCTTCACATTTTGTGCG.....ATGTCATCTAATAGTTTCAGAGAG 2349
839 leLeuAlaGlnAsnLeuHisAsnLeuValLysLysSerIleLeuAspLeu 855
2350 AGCTCCCTTCATCTATTTCCAAATTTAATCATCATTAATAATCTGATTT 2399
856 serGluAsnTyrLeuGluLysAsp.....GlyAsnGluAl 867
2400 GCGAGAAACAATCTGAGAGGAGCAATACCAATTTTGGCAATATTAG 2449
867 aLeuHisGluLeuIleAspArgMetAsnValLeuGlnLeuThrAlaL 884
2450 TAGCCTCAGGTTTGTGATATGAGAAAT.....AACAACTTCTGCG. 2492
884 euMetLeuProTrrpGlyCysAspValGlnGlySerLeuSerLeuLeu 900
```

```
2493 .. ACTCTTCCAACAATTTAGCATTTGATGTTCACTGATTAAGTCTCAAC 2540
901 LysHis.....LeuGluGluValPro..... 907
|||
2541 TTGCATTCGCAATGACATGACAGATGAATCCCTCGCTTTGGACAATTG 2590
908 ...GlnLeuValLysLeuGlyLeuLysAsnTrpArgLeuThrAspThr 922
|||
2591 CAAAAGCTGCAGATGTTGATTTAGAGACAAATCACTCAACGACACAT 2640
923 .....GluLeuArgIleLeuGlyAlaPhe 930
|||
2641 TTCCCATGTGTGGAACTTTGCCAGAGCTGAGATTTTAAGTTTGACA 2690
931 PheGlyLys.....AsnProLeuLys..... 937
|||
2691 TCGAATTAATTCATGACATGACCTTAAGATCATCAGGCGCTGAATCATGTT 2740
938 ...AsnPheGlnGlnLeuAsnLeuAlaGlyAsnArgValSerSerAspG 953
|||
2741 TCCGATCTCCGAATCATGATCTCTCTCGCATCATCTTCTCGCAAGAC 2789
953 LysTrpLeuAlaPheMetGlyValPheGluAsnLeuLysGlnLeuValPhe 969
|||
2790 .....TTACCAACGAGCTCTATTGAAACATTTGAAGGATGAGACA 2831
970 PheAspPheSerThrLysGlu.....PheLeuProAspPr 981
|||
2832 GTTGATTAACAAATGAGAACCAAGTATGAAAGCTTATGAGATGACTC 2881
981 AlaLeuVal.....ArgLysLeuSerGlnValLeu 992
|||
2882 GGTGCTATTGTGCACAAAGCATTTGAGCTTGAATTTGTGACAATCTGT 2931
992 eLysLeuThrPheLeu 997
|||
2932 CTTTGTACACAAATTATC 2948
seq_name: /cgrn2_6/protodata/2/lna/6b_COMB.seq:US-09-180-439-2
```

```
seq_documentation_block:
; Sequence 2, Application US/09180439
; Patent No. 6225532
; GENERAL INFORMATION:
; APPLICANT: Dixon, Mark S
; APPLICANT: Hatzixanthis, Kostas
; APPLICANT: Jones, David A
; TITLE OF INVENTION: Plant pathogen resistance genes and uses thereof
; FILE REFERENCE: 620 - 53
; CURRENT APPLICATION NUMBER: US/09/180,439
; EARLIER FILING DATE: 1998-12-06
; EARLIER APPLICATION NUMBER: PCT/GB97/01249
; EARLIER FILING DATE: 1997-05-08
; EARLIER APPLICATION NUMBER: GB 9609681.3
; EARLIER FILING DATE: 1996-05-09
; EARLIER APPLICATION NUMBER: GB 9619924.5
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 3979
; TYPE: DNA
; ORGANISM: Lycopersicon esculentum
US-09-180-439-2

alignment_scores:
Quality: 172.00 Length: 639
Ratio: 0.533 Gaps: 38
Percent Similarity: 50.348 Percent Identity: 23.005

alignment_block:
```

```
US-09-697-089-2 x US-09-180-439-2
Align seg 1/1 to: US-09-180-439-2 from: 1 to: 3979
459 LeuThrSerHisGluProGlnGluValThrLysGlyAsnGlyThrLeuG 475
|||
1260 CTTTCGGCTTATTCTCGAAGAAATA.....GGTACTCTAG 1297
475 nLysMetValSerIleSer..AspIleThrSerThrThrSerLeu 491
|||
1298 GTCCTTACTAGCTATCTTGATATCAAC..... 1328
491 euArgTyrThrCysGlySerSerValGluAlaThrArgAlaValMetLys 507
|||
1329 .....TTTCTTAGTGTGTTCCATTCCCTGCTTCAATGGGGAATCTGAAC 1370
508 HisLeuAlaAlaValTyrGlnHisGlyCysLeuLeuGlyLeuSerIleAl 524
|||
1371 AACTGTCTTTTGTATCTTACATATATACAGCTTCTGCTCTATT.. 1418
524 aLysArgProLeuTrpArgGlnGluSerLeuGlnSerValLysAsnThr 541
|||
1419 .....CTGAAAGAAATAGTTTACCTTAAGTCACTTA 1449
541 hrGluGlnGluLeuLysAlaIleAsnIleAsnSerPheValGluCys 557
|||
1450 CTAAAG.....CTATCTTGG 1463
558 GlyIleHisLeuTyrGlnGluSerThrSerLysSerAla..... 570
|||
1464 GGTATCAACTTTCTAGAGGTTCATCCCTGCTTATGGGGAATCTMAA 1513
571 ...LeuSerGlnGluPheGlnAlaPhePheGlnGlySerLeuTyrI 586
|||
1514 CAACTTGTCTAG.....TTGATCTTTTACA 1539
586 Leasn.....SerGlyAsnIleProasp.....TyrLeu..... 595
|||
1540 ATAATAGCTTTCTGCTCTATCTCTGAAGAAATAGTTTACTTAAGCTCT 1589
596 .....PheAspPhePheGluHis.....LeuProAs 604
|||
1590 CTTACTTACTAGATTGGTGAGAAATGCTCTTATGCTCTATCTCTTC 1639
604 nCysAlaSerAlaLeuAspPheIle..LysLeuAspPheTyrGlyGlyAla 620
|||
1640 TTCATTTGGGGAATCTTAACACTGTCTAGTTGGATCTTTACAAATAATA 1689
620 lAmetAlaSerTrpGluLysAlaAlaGluAspThrGlyIleHisMet 636
|||
1690 AGCTTCTGCG.....TCTATCTCTGAAGAAATAGTTTACTTAAGCTCT 1733
637 .....GlnGluAlaProGluThrTyrIleProse 646
|||
1734 CTTACTTACTAGATTGGTGAGAAATGCTCTTATGCTCTATCTCTGCG 1783
646 rArgAlaValSerLeuPheAsnTrpLysGlnGluPheArgThrLeuG 663
|||
1784 TTCATTTGGGGAATCTG.....AACACTTGTATGATGTGT 1818
663 lValThrLeuArgAspPheSerLysLeuAsnLysGlnAspIleThrTyr 679
|||
1819 ATCTTTACATATATGCTTCTGCTCTATCTCTGAAGAAATAGCTTAC 1868
680 LeuGlyLysIlePheSerSerAlaThrSerLeuArgLeuGlnIleLysAr 696
|||
1869 CTG.....AGTTCTCTTACTGACATCATATTTGGGAATATAC.. 1904
696 gCysAlaGlyValAlaGlySerLeuSerLeuValLeuSerThrCysLysAla 713
|||
1905 .....TCTCTTAATGCTCTATCTCTGCTTCAATTTGGGGAATCTGAACA 1947
713 snIleTyrSerLeuMetValGluAlaSerProLeuThr.....Ile 726
```

```

1948 ACTGTGTTATGTTGTAATCTTACAAATATACGCTTTCTGGCTCTATTCT 1997
727 GUAAPGLUARHGISLethrSerValThrasnLeuylThrLeuSeril 743
1998 GAAGAATAGGTACCTGAGTCTCTACTGAACTA..... 2033
743 eHISAPLeuGlnAsnGlnArgLeuProGlyGlyLeuThraspSerLeuG 760
2034 ....TTTTGGTAATAACTCTCTAATGGCTCTATTCTCGCTCATTTG 2079
760 lYAsnLeuLysAsnLeuThrLysLeuileMeLaspAsnileLysMetAsn 776
2080 GGAAATTAACAACCTGTCTAGTGTATCTTACAAATATACAGCTTTCT 2129
777 GlnGluAspAlaileLysLeuAlaGly...LeuLysAsnLeuLysL 792
2130 ....GGCTCTATTCCTGCTCATTTGGCAATATAGAAATCTGCAAAC 2173
792 sMetCysLeuPheHISLeuThrHISLeuSerAsp.....lLeGlyG 806
2174 TCTG.....TTTCTCAGTGATATACGATCTCATTTGGG 2205
806 lueGlyMetAspTyrileValLysSerLeuSerSerGlu..... 818
2206 AAATTCCTTCATTTGTGTGCAATTTGACATCACTGCAAGCTGTATATG 2255
819 .....ProCys.....AspLeuG1 823
2256 TCGAGAAACAATTTGAAGGAAAGTTCCGCAATGTTGGTAATATCAG 2305
823 uGluileGlnLeuValSerCysCysLeuSerAlaAsnAlaValLys...I 839
2306 TGACCTTCACATTTTGTG....ATGTCATCTAATATGTTTCAGAGGAG 2349
839 lLeuValGlnAsnLeuHISAsnLeuValLysLeuSerileLeuAspLeu 855
2350 AGTCCCTTCATCTATTTCCAATTTAACATCACTAAATACCTTGATTT 2399
856 SerGluAsnTyrLeuGluLysAsp.....GlyAsnGluAl 867
2400 GGCAGAAACAATCTGGAGGAGCAATACACAATTTTGGCAATATTAG 2449
867 AlauHISGlnLeuileAspArgMetAsnValLeuGlnGluLeuThral 884
2450 TACCTTCACAGTTTGTGATATGCAAT.....AACAACTTTCCTGGG. 2492
884 eueMetLeuProTrrGlyCysAspValGlnGlySerLeuSerSerLeu 900
2493 ..ACTCTTCACAACAATTTAGCATTTGATGATGATAGTCGAC 2540
901 LysHIS.....LeuGlnGluValPro..... 907
2541 TTCATTCGCAATGATGACATGAGAAATCCCTCGCTTTGGACAATTG 2590
908 ....GlnLeuValLysLeuGlyLeuLysAsnTrrPargLeuThraspHr. 922
2591 CAAAAGCTGCAAGTCTCTGATTTAGGAGACAATCAACGACGACAT 2640
923 .....GluileArgileLeuGlyAlaPhe 930
2641 TTCGCATGTGTGGAACTTTGCCAGAGCTGAGAGTTTAAGGTGACA 2690
931 PheGlyLys.....AsnProLeuLys..... 937
2691 TCGAATAAATTTGCATGACCTATAGATCATCAAGGGCTGAATCATGTT 2740
938 ....AsnPheGlnGlnLeuAsnLeuAlaGlyAsnArgValSerSerAsp 953
2741 TCTGTATCTCGCAATCATGATCTCTCCCAATGCAATTCCTCGCAAC 2789
953 LyrTrpLeuAlaPheMetGlyValPheGlnAsnLeuLysGlnLeuValPhe 969

```

```

2790 .....TTACCAAGAGTCTATTGACATTTGAAGGAGTACAGACA 2831
970 PheAspPheSerThrLysGlu.....PheLeuProAspPr 981
2832 GTTGATTAACAATGAGAGCAACAGATTATGAAGCACTATACGATGACTC 2881
981 oAlaLeuVal.....ArgLysLeuSerGlnValLeuS 992
2882 GGTGGTAGTTGTGACAAAGGATTTGAGCTTGAAATTTGTAGAAATTCGT 2931
992 eRlyLysLeuThrPheLeu 997
2932 CTTTGTACACAATTATC 2948

seq_name: /cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-353-585-4

seq_documentation_block:
; Sequence 4, Application US/09353585
; Patent No. 6287865
; GENERAL INFORMATION:
; APPLICANT: Dixon, Mark S
; Jones, David A
; Jones, Jonathan DG
; TITLE OF INVENTION: Plant pathogen resistance genes and uses thereof
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Nixon & Vanderhye PC
; STREET: 8th Floor, 1100 No. 6287865th Glebe Road
; CITY: Arlington
; STATE: Virginia
; COUNTRY: United States of America
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/353,585
; FILING DATE: 15-Jul-1999
; CLASSIFICATION: C12N 15/29, 15/82, A01H 5/00, A01N 65/00, C12Q 1/68

PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/930,277
; FILING DATE: 27-OCT-1997
; APPLICATION NUMBER: PCT/GB96/00785
; FILING DATE: 01-APR-1996
; APPLICATION NUMBER: GB 9506658.5
; FILING DATE: 31-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Ms Mary J Wilson
; REGISTRATION NUMBER: 32,955
; REFERENCE/DOCKET NUMBER: 620-69
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3573 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Tomato
; STRAIN: Cf2
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-353-585-4

```

alignment_scores: Length: 645
 Quality: 165.00
 Ratio: 0.546
 Gaps: 31
 Percent Similarity: 46.822 Percent Identity: 21.395

alignment_block:
 US-09-697-089-2 x US-09-353-585-4 ..

Align seg 1/1 to: US-09-353-585-4 from: 1 to: 3573

```

425 GlyLeuLeuCysLysTyrThrAlaGlnArgPheLysProLysTyrLysPhe 441
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
440 GGTTCCTAGCCAGCTCAGATCATCCGC .....AT 441
441 ephenHsLysSerPheGlnGluTyrThrAlaGlnArgLysSerLys 458
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
442 ATTTCACAAATCAA ..... 454
458 euleuTherHsGlnProGlnGluValThrLysGlnGlyTyrLeu 474
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
455 ..TTAAATGATTTATTCCTAAAGAAATA .....GGTTCCTTA 490
475 GlnLysMetValSerLysSerAspIleThrSerThrTyrSerSerLeu 491
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
491 AGGTCTCTACTAAGCTATCT...TTGGATACAACTTCTTAGT ..... 532
491 uArgGlyThrCysGlySerSerValGlnAlaThrArgAlaValMetLys 508
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
533 .....GGTTCATTCCTCTCAGTGGGGAATCTGMAA 566
508 IsLeuAlaAlaValTyrGlnHsGlyCysLeuGlnLysLeuSerIleAla 524
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
567 ACTGCTCTTTTGTATCTTTACAAATATACAGCTTCTGCTCTAT... 613
525 LysArgProLeuThrArgGlnGlnSerLeuGlnSerValLysAsnThr 541
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
614 .....CCTGAAGAAATAGATTACCTAAGATCTGTAC 645
541 rGlnGlnGluIleLeuLysAlaIleAsnIleAsnSerPheValGlnCys 558
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
646 TGAAGTAGATTTG ..... 658
558 LysIleHsLeuTyrGlnGlnSerThrSerLysSerAlaLeuSerGln 574
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
659 .....AGTGATATGCTCTTAATGCGCTCT 682
575 PheGlnAla .....PhePheGln 580
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
683 ATTCCGCTTCATTTGGGAATATGACAACTGTCTTTTGTCTTTTA 732
580 nGlyLysSerLeuTyrIleAsnSerGlnLysIleProAsp .....T 594
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
733 TGGAAATCAGCTT .....TCTGGCTATTCCTGAAGAAATATGTT 773
594 yrLeu .....PheAspPheGlnHsIleLeuProAsnCys 605
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
774 ACCTAAGATCTCTACTTACCTAGATTTAGTGAGAAATGCTCTTAATG 823
606 Ala .....SerAlaLeuAspPheIleLysLe 614
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
824 TCTATTCCTGCTTCATTTGGGAATTTGACAACTGTCTTTTGTCTTCT 873
614 uAspPheTyrGlnGlyAlaMetAlaSerTrpGlnLysAlaIleGlnAsp 631
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
874 T .....TATGGAATCAGCTTCTGCG .....TCTATTCCTGAAGAA 911
631 hGlnGlyIleHsMetGlnGlnAlaProGlnThrTyrIleProSerArg 647
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
912 TAGGT .....TACCTA ..... 922
648 AlaValSerLeuPheAsnTrpLysGlnGlnPheArgThrLeuGlnVal 664
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
923 .....AGATCTCTTAATGT 936
```

```

664 LThrLeuArgAspPheSerLysLeuAsnLysGlnAspIleThrTyrLeu 681
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
937 CCTAGCTTTGAGTGAGAAATGCTCTTAAT ..... 964
681 LysIlePheSerSerAlaThrSerLeuArgGlnIleLysArgCys 697
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
964 ..... 964
698 AlaGlyValAlaGlySerLeuSerLeuValLeuSerThrCysLysAsn 714
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
965 .....GGCTATATCTCTGCTTATCTGGAATCTGAAATCTGAAACCT 1002
714 eTyrSerLeuMetValGlnAlaSerProLeuThrIleGlnAspGlnArg 731
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1003 GTCTAGTTGAATCTTGTAAATATCAGCTTCTGCTCTATTCCT... 1048
731 IsIleThrSerValThrAsnLeuLysThrLeuSerIleHsAspLeuGln 747
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1049 .....GCTTCATTTGGGAATCTGACAACTGTCTATGTTGTCTTTAC 1093
748 AsnGlnArgLeuProGlnLysLeuThrAspSerLeuGlnLysLysAs 764
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1094 AATACACACTTCTGCTCTATCTCTGCTTATCTGGAATCTGAAACAA 1143
764 PleuThrLysLeuIleMetAspAsnIleLysMetAsnGlnGlnAspAla 781
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1144 CTGTCTATGTTGTATCTTACAAATATCAGCTTCTGCTCTATTCCTG 1193
781 LeuLysLeuAlaGlnGlyLeuLysAsnLeuLysLysMetCysLeuPhe 797
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1194 CTTCATTTGGGG...AATCTGAACAACTTCTATGAGTTGATCTTACAA 1240
798 LeuThrHsLeuSerAspIleGlnGlnMetAspTyrIleValLys... 813
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1241 AATACACTTCTGCTCTATCTCTGGAAGAAATAGTACTGAGTCTCT 1290
814 .....SerLeuSerSerGlnProCysAsp ..... 821
1291 TACTATCTAGATTTGATTAATATCAGTAAATGAGTTAATTCCTGCT 1340
822 .....LeuGlnGluIleGln 826
1341 CATTTGCGAATATGAGCACTGCGCTTTTGTGTTTCTTATGAAATCAG 1390
827 LeuValSerCysCysLeuSerAlaAsnAlaValLysIleLeuAlaGln 843
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1391 CTTCGTAAGCTCT .....GTTCCGTGAAGA 1413
843 nLeuHsAsnLeuValLysLeuSerIleLeuAspLeuSerGlnAsnTyr 860
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1414 AATAGGTTCAGTAAAGCTCTTAATGCTCTGATTTGATGAGAAATGCT 1463
860 euGlnLysAspLysGlnAlaIleLeuHsGlnLeuIleAspLysMetAsn 876
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1464 TT .....AATGCTCTATTCCTGCTCTATTCGAGCAATTTGAAAC 1501
877 ValLeuGlnGlnLeuThrAlaLeuMetLeuProTrpGlnCysAspAla 893
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1502 AACTTCCTAGTGTGAATCTT .....GTAA 1527
893 nGlySerLeuSerSerLeuLeuLysHsLeuGlnGlnValProGlnLeu 910
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1528 TAATCAGCTTCTGCTCTATTCCT .....GAGGAAATAGCTTACCTA 1571
910 aLysLeuGlnLysLeuAsnTrpArgLeuThrAspHsGlnIleArg... 925
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1572 GGTCTCTTAATGCTCTGAT .....TTGAGTGAGAAATGCTCTTAATG 1615
926 IleLeuGlyAlaPhePheGlnLysAsnProLeuLysAsnPheGlnGln 942
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1616 TCTATTCCTGCTTCATTCGGAAT .....TTGACAACTTCTAGAGTT 1659
```



```

2945 TCATATTGAACATTGAAAGGATGAGACAGTGTGATAAACAATGAGAG 2994
976 lu.....PheLeuProAspProAlaLeuVal..... 984
2995 AACCAAGTATGAAAGCTATTACGATGACGCGGTGAGTGTGACAAAG 3044
985 .....ArgLysLeuSerGlnValLeuSerLysLeuThrPheLeu 997
3045 GGATTGGAGCTTGAAATGTGACAAATTCCTCTTTGTACACAAATTATC 3092
seq_name: /cgn2_6/ptodata/2/1na/5A_COMB.seq: us-08-353-700-2

```

```

seq_documentation_block:
Sequence 2, Application US/08353700
Patent No. 5599919
GENERAL INFORMATION:
APPLICANT: YEN, TIMOTHY J.
APPLICANT: RATNER, JEROME B.
TITLE OF INVENTION: NUCLEIC ACID ENCODING A
TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
TITLE OF INVENTION: AND METHODS OF USE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: DANN, DOREMAN, HERRELL AND SKILLMAN
STREET: 1601 MARKET STREET, SUITE 720
CITY: PHILADELPHIA
STATE: PA
COUNTRY: USA
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/353,700
FILING DATE: 09-DEC-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: REED, JANET E.
REGISTRATION NUMBER: 36,252
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 10136 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHEetical: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: HUMAN
us-08-353-700-2

```

```

alignment_scores:
Quality: 149.50 Length: 1156
Ratio: 0.266 Caps: 52
Percent Similarity: 48.702 Percent Identity: 18.426

```

alignment_block:

```

us-09-697-089-2 x us-08-353-700-2 ..
Align seg 1/1 to: us-08-353-700-2 from: 1 to: 10136

```

```

39 AsnIleIleCysGlyValGlnGlnAspAlaAlaArgGlyIleIle 55
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2220 AACGTGTGAGACAGTAAGTCAGTACGAGGAGTAAGACACCGAGAAA...CTAGC 2266

```

```

55 eHisMetIleLeuLysLysGlySerGlnSerCys..... 66
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2267 TTAATATGAGAGCTACAGCAACAAAGCTGATCTCAGATCGAAACATCAGA 2316
67 .....AsnLeuPheLeuLysSerLeuLysGlnIlePro 79
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2317 AGCAATATAGAAAATATGCTTTGAAGACT..... 2345
80 LeuPheGlnAspLeuAsnGlyGln.....SerLeuPheHis.....G1 92
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2346 .....TCTCAGCTTACTGCGCAAGTGAAGATCTGAAGACCAAGCTTCA 2389
92 nThrSerGlnGlyAspLeuAspLeuAlaGlnAspLeuLysAspLeu 109
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2390 GTTACTGTCAAAATGAATATATGACACAAAGACCGGTGTTACCAAGCTTGC 2439
109 yHisThrProSerPheLeuAsnPheTyProLeuGlyGlnAspLeu 125
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2440 ATGCCGATATGAGAGCCTCAGGAGATCTGCTAAATCCAAAGATGCTTCT 2489
126 IleIlePheAsn.....LeuLysSerThrPheThrG1 136
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2490 GTGGTACAAATGAAGATCATCAGACAGAGCTTTGGCTTTGATCAGCA 2539
136 uProValIleuTrpArgLysAspGlnHisHisArgValGlnIleu 153
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2540 GCCTGCCATG.....CATCATTCCTTTGCAATATATATG 2574
153 hLeuAsnGlyLeuLeuGlnAlaLeuGlnSerProCysIleIleGly 169
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2575 GAGACAAAGCAAGCATGCTTCAGAGAGAGATGATGCTGTTAAGCA 2624
170 GluSerGlyLysGlyLysSerThrLeuGln...ArgIleAlaMetIle 185
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2625 GACCAAGTCCGAAAATTTGCCATCTCAAAATAGAGTTGATTCAT 2674
185 uTrpGlySerGlyLysCysLysAlaLeuThrLysPheLysPheValPhe 202
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2675 T.....GAATTTT 2682
202 hLeuArgLeuSerArgAlaGlnGlyLysLeuPheGlnThrLeuCysAsp 218
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2683 CATTAAGATCTCAAAAACAGATGATGACTGACCTCAAAAGCAGTGTGA 2732
219 GlnLeuLeuAspLysProGlyThrIleArgLysGlnThrPheMetAlaMet 235
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2733 GAGTTGGTCAAAATCAAAAGAGAAATAGAAAGAAAT..... 2768
235 tLeuLeuLysLeuArgGlnArgValLeuPheLeuLeuAspGlyTyAsnG 252
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2769 .CTCATGAAGACGACAGACAGATG.....CATCAAA 2796
252 hLeuPheLysProGlnAsnGlyProGlnIleGlnAlaLeuIleLysGlnAsn 268
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2797 GTTTGTGCTGCAAAACAGTACGCGCATTAAGTTACAGAGACAGACT 2846
269 HisArgPheLysAsnMetValIleValIleThrThrThrGlnCysLeuArg 285
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2847 TCTGCTCACCAGAAAT...GTGTGCTGAACCTTAAGTCCCTTGAGAA 2893
285 gHisIleArgGlnPheGlyAlaLeuThrAlaGlnValGlyAspMetThrG 302
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2894 CAAGCAAAAAGAGCTGCAACTTTAAATATATAGCTA..... 2930
302 hAspSerAlaGlnAlaLeuIleArgGlnVal...LeuIleLysGlnLeu 317
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2931 ..GAAACTGACGACGACAGATTAAGAAATTAAGAAAGCAACCATCTA 2978
318 AlaGlnGlyLeuLeuLeuGlnIleGlnLysSerArgCysLeuArgAsnLe 334
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2979 CTTGAAGACTCTCTAAAGGAGCTCAACAACCTTTATCCGAACCCCTAAGCTT 3028
334 uMetLysThrProLeuPheValValIleThrCysAlaIleGlnMetGlyG 351

```

```

3029 GGAGAGGAAAGAAATGAGTTCATCATTTCT.....TTAAATA 3066
351 luserGluPheHisSerHisThrInThr.....ThrLeuPheHisThr 365
3067 AAAGGAAATTTGAAGAGCTGACCCAGAGAAATGGACTCTTAAGAAATT 3116
366 PheTyrAspLeu.....LeuIleGlnLysAsnLys.. 375
3117 AATGATCCTTAATCAAGAGAGTGAATCAATCCAGAAAGCTGACAG 3166
376 .....HisLysHisLysGlyAlaAla...AlaSerA 385
3167 TTTTGCAACTATATAGATGAAGGAGAAAGCATTTTCAGATTATCTG 3216
385 sspHeiLeaArgSer.....LeuAspHisCysGlyAspLeu 396
3217 ATCAGTACACCAAGAAAACCTTATTTTACTACAAAGATGGAAGAAACC 3266
397 .....AlaLeuGlnGlyAlaPheSerHisLysPheAspPheGlnLeuG1 411
3267 GGAATGCAATATGAGATCTT...AGTCAAAATATACAAA...GCAGCACA 3310
411 nasPValSerSerValAsnGlnAspValLeuLeuThrThrGlnLeuEuc 428
3311 GGAAGAAGATTCTAAATTAGATGCTGCTAAATGAATGCACATAGTCTTT 3360
428 yslLysTyrThrAlaGlnArgPheLysProLysTyrLysPhePheHisLys 444
3361 GTGAAATAGCAAAATAGATTGGACACAGCTAAAGAAAGCATTTTGCAAG 3410
445 SerPheGlnGlyTyr.....ThrAlaGlyArgArgLeuSerSerLeuL 449
3411 GAACACCAAGAAATCTTAAACAAATATGACATTTGCTGAGAAAGAAATCA 3460
450 .....ThrAlaGlyArgArgLeuSerSerLeuL 459
3461 GAATCTGATGCTAGATTGAGACAGCTGCAGCAAGCTCTGAGATCTGAGA 3510
459 eutHrSerHisGluProGlnGlnValThrLysGlyAsnGlyTyrLeuGln 475
3511 TGACAGATAAACCAACAAATCTTAAGACGAGCGTGTGTTAAAGCAA 3560
476 LysMetValSerLysSerAsp.....IleHrSerThrTyrSe 488
3561 GAAATCAATGACTTTAAAGGAAACAAACAAATGCCAAAGGAAGTAA 3610
488 rSerLeuLeuArgTyrThrCysGlySerSerValGlnAlaThrArgAla 505
3611 TGACTTATTTACAAAGAGAT.....GAACAGCTGATGAAGG 3645
505 alMetLysHisLeuAlaAlaValTyrGlnHisGlyLysLeuLeuGlyLeu 521
3646 TAATGAG.....ACTAACATGATGT..... 3668
522 SerIleAlaLysArgProLeuTyrPArgGlnLysLeuGlnSer.... 536
3669 .....CAAATCTAGATCAAGAAC 3688
537 ValLysAsnThrThrGlnGlnGlnIleLeuLysAlaIleAsnIleAsn 553
3689 AATTAGGAACCTGTGAAAGAAAGAGAGAGTGAAGAAATCAATGTAATT 3738
553 ePheValGlnCysGlyIleHisLeuTyrGlnLysSer...ThrSerLys 568
3739 TTTAAACCTCAATGATCTTGAAGTTAAAGAAATTTCTCTAGATAGTTAT 3788
569 SerAlaLeuSerGlnGlnPheGlnAlaPhePheGlnGlyLysSerLeuTy 585
3789 AATCGCGAGTTGTGCAATTAGAAGCTATGCTAAGAAATTAAGAAATTAA 3838
585 rIleAsnSerLysAsnIleProAspTyrLeuPheAspPhePheGlnHisL 602

```

```

3839 ACTTCAGGAAAGT.....GAGAGG 3858
602 euProAsnCysAlaSer...AlaLeuAspPheIleLysLeuAspPheTyr 617
3859 AGAAGAGAGTGGCTGCACACATTAATTACAGCAATTAGAGGAGATCTGAA 3908
618 GlyLysAlaMetAlaSerTyrGlnLysAlaAlaGlnAspThrGlyGlyI 634
3909 ACCAGCAATTTGCAAGACATGCAG.....TCACAGAAATTAAGTGCCT 3952
634 eHisMetGlnGlu.....AlaProGlnThrTyrIleProSer...Arg 648
3953 TTAAGACTGTGAATATAGATGCGAAGAAAGATATATTTCAAGGCTCATG 4002
648 lavalSerLeuPhePheAsnTyrPlyGlnGlnPheArg...ThrLeuGln 663
4003 AGTGTCTACACAGTCAACAAACGCAATGACACCTTCAGTGTCTCTGCA 4052
664 ValThrLeuArgAspPheSerLysLeuAsnLysGlnAspIleThrTyrLe 680
4053 ACAACAATGACACAGCTGAATGAGCTAGAGAAATATGTGAATATCTGCA 4102
680 uGlyLysIlePheSerSerAlaThrSerLeuArgLeuGlnIleLysArgC 697
4103 GGTGAAAGATATGAACTCTGTAAGCTGAAGTGAATGATCAAGCTCAAT 4152
697 yslAlaGlyValAlaGlySerLeuSerLeuValLeuSerThrCysLysAsn 713
4153 GTATCACACGCACTAGCAAAATGGCAGAGAGTGAAGGAAACTCTAAAT 4202
714 IleTyrSerLeuMetValGlnAlaSerProLeu.....ThrI1 726
4203 GAAGTTAAATATTAATATGATGACAGCTGCTCTCCATGGTGAATTAG 4252
726 eGluAsp.....GluArgHisI 732
4253 GGAAGACATPACCAGAGGTGAATTTGGTGAACAACAAATGAACAGCAC 4302
732 lenthSerValThrAsnLeu.....LysThrLeuSer 742
4303 CTGTGCTTTGGTCCATCTGACAGAGATATPCTAGACACTTGCA 4352
743 IleHisAspLeuGlnAsnGlnArgLeuProGlyGlyLeuThrAspSerLe 759
4353 TTGTCAGACAAAGAAATTCAATGACACTTTCGCAATTTGCAAGAAATT 4402
759 uGlyAsnLeuLysAsnLeuThrLysLeuIleMetAsp..... 771
4403 CTATCTTTACAAAGTGAACACAAATTTTACATGATCAGCAGCTGCA 4452
771 ..... 771
4453 TGAGCTTAATATGTCAGAGCTGCAGACCTATGTTGACTATTAAGGCC 4502
772 .....AsnIleLysMetAsnGlnGlnAspAlaI1 781
4503 GAAATTTTGTCTTGTCAACCAATCTGAGAACTTCAAGGTGACTGTGT 4552
781 eLys.....LeuAlaGlnGlyLeu...LysAsnLeuLysL 792
4553 GAAGGAGATGAGCTGGCTTGAGAGAGGGCTGCTTCATCCCTGTGCAT 4602
792 ysmetCysLeuPheHisLeuThrHisLeuSerAspIleGlyGlnGlyMet 808
4603 CCTCTGTGTGCTGCAGACGCTCTAGTCTTACATTTGGGAGACTCCTCC 810
809 AspTyr..... 810
4653 TTTTACAGAGCTTTTGAACAGACAGAGATATGCTCTTTGAGTAA 4702
811 lIleValLysSerLeuSerSerGluProCysAspLeuGlnGlnIleGlnL 827
4703 TTTAGAGAGGGCTGTTTCAGCAACAGTGCAGTGTAGATGAATATTTT 4752

```

```

827 euValserCysCysLeuSerAlaAlaVallys.....Ile 839
      |||
4753 GCAGCAGTCTGCAGACCTATGTTGACTCATTAAGGCCGAATTTGGTC 4802
      |||
840 LeuAlaGlnAsnLeuHisAsn.....LeuVallysLeuSerIle 852
      |||
4803 TTGTCAACGATCTGAGAACTTTCAAGGTGACTGGTGAAGAGATGCA 4852
      |||
852 eLeuAspLeuSerGluAsnTyrLeu..... 860
      |||
4853 GCTGGGCTTGGAGAGGGCTGCTCCATCCCTGTCATCCCTTGTGGTC 4902
      |||
861 .....GluLysAspGluAsnGluAla 867
      |||
4903 CTCACAGCTCTAGCTTAGACAGTTTGGAGACATCCCTTTTACAGAGCT 4952
      |||
868 LeuHisGluLeuIleAspArgMetAsnValLeuGluGlnLeuThrAlaLe 884
      |||
4953 CTTTGTAGAACACAGACGAGATATGCTCTTTGTAGTAATTTAGAACGGGT 5002
      |||
884 uMetLeuProTyrGlyCysAspArgIleGlnLysLeuSerSerLeuLeu 901
      |||
5003 TGTTCACGAAACCACTGACAGTGTAGATGATATTTCACAGCTCTGC 5052
      |||
901 yshis.....LeuGluGluValProGlnLeuValLysLeuGly 913
      |||
5053 AGCAGCAGATCTGACCGAAGAAACCCCTTCGGCCCGCAGGAGGCT 5102
      |||
914 LeuLysAsnTyrIleuThrAspThrGluIleArgIleLeuGlyAlaPh 930
      |||
5103 GTTGAA.....GAGCTTGAGTCCCTCTGTGAGGT 5131
      |||
930 ePheGlyLysAsnProLeuLysAsnPheGlnGlnLeuAsnLeuAlaGlyA 947
      |||
5132 GTACCG.....CAGTCCCTCGAGAGAGCTAGAG 5160
      |||
947 snArgValSerSerAspGlyTyrPheAlaPheMetGlyValPheGluAsn 963
      |||
5161 AGAAATGGAAGTCAAGGATTTATGAAAATTAAGCAATTT..CAAGAG 5207
      |||
964 LeuLysGlnLeuValPhe.....PheAspPheSerThrly 975
      |||
5208 CTCGACACCTTATTAAGTTCTGAAGCAAGAGCTGACTGCTTAGAA 5257
      |||
975 sGluPheLeuProAspProAlaLeuValArgLysLeuSerGlnValLeu 992
      |||
5258 GCAGTATTTGTCAGAAAT.....GACAGTGGCAACAGAACCTGA 5298
      |||
992 eLysLeuThrPheLeuGlnGluAlaArgLeuValGlyTyrGlnPheAsp 1008
      |||
5299 CAAGCTGACTCTGAGATGAGACTCAAGTTGGCGCAGAAAAAGAACAG 5348
      |||
1009 AspAspAspLeuSerVal 1014
      |||
5349 ACGGAACAACCTGCACTT 5366

```

seq_name: /cgn2_6/ptodata/2/ina/PCUS.COMB.seq:PCT-US95-16216-2

```

seq_documentation_block:
: Sequence 2, Application PC/RTUS9516216
: GENERAL INFORMATION:
: APPLICANT: Yen, Timothy J.
: APPLICANT: Ratner, Jerome B.
: TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESS: Dann, Dorfman, Herrell and Skillman
: STREET: 1601 Market Street Suite 720
: CITY: Philadelphia
: STATE: PA
: COUNTRY: USA

```

```

: ZIP: 19103-2307
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentln Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/16216
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/353,700
: FILING DATE: 09-DEC-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Reed, Janet E.
: REGISTRATION NUMBER: 36,252
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (215) 563-4100
: TELEFAX: (215) 563-4044
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 10136 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: PCT-US95-16216-2

```

```

alignment_scores:
: Quality: 149.50 Length: 1156
: Ratio: 0.266 Gaps: 52
: Percent Similarity: 48.702 Percent Identity: 18.426

```

alignment_block:
US-09-697-089-2 x PCT-US95-16216-2 ..

Align seg 1/1 to: PCT-US95-16216-2 from: 1 to: 10136

```

39 AsnIleIleCysCysGluLysValGluGlnAspAlaAlaArgGlyIleI 55
      |||
2220 AACGCTTGACAGTAAGTCACTGAGAGTGAAGACCCAGAA...CTAGC 2266
      |||
55 eHisMetIleuLysLysGlySerGluSerCys..... 66
      |||
2267 TTATATGAGACTACAGCAAGAGCTGAGTTCTACATCAGAAACATCAGA 2316
      |||
67 .....AsnLeuPheLysSerLeuLysGluTyrAsnTyrPro 79
      |||
2317 AGCAATTAAGAAATATGTGTTGAAGACT..... 2345
      |||
80 LeuPheGlnAspLeuAsnGlyln.....SerLeuPheHis.....G1 92
      |||
2346 .....TCTCAGCTTACTGGCAGTGAAGATCTTAAGAACACACTTCA 2389
      |||
92 nThrSerGluGlyAspLeuAspAspLeuAlaGlnAspLeuLysAspLeu 109
      |||
2390 GTTACTGTCAAAATGAATTAATGACAAAGCCGGTGTACCAAGACTTGC 2439
      |||
109 yHisThrProSerPheLeuAsnPheTyrProLeuGlyGluAspLeu 125
      |||
2440 ATGCCGAATATAGAGACCTCAGGAGATCTGCTAAATCCAAAGATCTTCT 2489
      |||
126 IleIlePheAsn.....LeuLysSerThrPheThrG1 136
      |||
2490 CTGGGACAAATGAGATCATCAGAGAGTCTTTGGCTTTGATCAGCA 2539
      |||
136 uProValLeuTyrArgLysAspGlnHisHisArgValGluGlnLeu 153
      |||
2540 GCCTGCCATG.....CATCATTCCTTTGCAATATATATG 2574

```



```

4153 GATTCACAGCACTAGAGAAAATGGCAGAAAGGTAGGAAACTACTAAT 4202
714 ILeTyrSerLeuMetValGluAlaSerProLeu.....Thr11 726
4203 GAAGTTAAATATATTAATGATGACAGTGGCTTCTCCATGCTAGTTAG 4252
726 eGluAsp.....GluArgHis1 732
4253 GGAAGACATACACGAGAGTGAATTTGGTGACACCAACCAATGAAACACACC 4302
732 LeThrsValThrsLeu.....LysThrsLeuSer 742
4303 CTGTGCTTTGGCTCCATTGGACGAGAGTAATTCCTACGACGACTGTGACA 4352
743 ILeHisAspLeuGlnAsnGlnArgLeuProGlyGlyLeuThrsSerLe 759
4353 TTGTTCAGACAAAGAGACTCAATGCACTTTGGCCGAATTCAGAGAAAT 4402
759 uGlyAsnLeuLysAsnLeuThrsLeuIleMetAsp..... 771
4403 CTATCTTTACAAAGTGAAACAAATTTTACATGATCAGCACTGTGACA 4452
771 ..... 771
4453 TGAGCTCTAAATGTGACAGCTGCAGACCTATGCTACTATTAAAGGCC 4502
772 .....AsnIleLysMetAsnGluGluAspAla11 781
4503 GAAATTTGGTCTTGTCAACGAATCTGAGAAACTTTCAAGGTGACTTGGT 4552
781 eLys.....LeuAlaGluGlyLeu..LysAsnLeuLysL 792
4553 GAAGGAGATGCACCTGGCTTGAGAGAGGGCTCGTTCATCCTCTCAT 4602
792 ySMeLcysLeuPheHisLeuThrsHisLeuSerAspIleGlyGlyMet 808
4603 CCTCTTGTGCTGCACAGCTGACTTGTAGCAGTTGGAGACTCTCTCC 4652
809 AspTyr..... 810
4653 TTTTACAGAGCTTTTACACAGACAGAGATATCTCTTTTGAGTAA 4702
811 ILeValLysSerLeuSerSerGluProCysAspLeuGluGluLeuL 827
4703 TTTAGAAAGGGCTGTTTCAGCAACAGCTGAGTGAAGTATTTT 4752
827 euValSerCysLeuSerAlaAsnAlaValLys.....Ile 839
4753 GCAGCAGCTCTGCAGACCTATGTTGACTCTTAAAGGCCGAAATTTGGTC 4802
840 LeuAlaGlnAsnLeuHisAsn.....LeuValLysLeuSer11 852
4803 TTGTCAACGAATCTGAGAAACTTTCAAGGTGACTTGTGTAAGAGATGCA 4852
852 eLeuAspLeuSerGluAsnTyrLeu..... 860
4853 GCTGGGCTTGAGAGAGGGCTCGTTCCATCCCTGTATCCTCTGTGTGC 4902
861 .....GluLysAspGlyAsnGluAla 867
4903 CTCACAGCTCTAGTCTTACAGATTGGAGACTCCCTCTTACAGAGCT 4952
868 LeuHisGluLeuIleAspArgMetAsnValLeuGluGlnLeuThrsAla 884
4953 CTTTGAACACAGACAGAGATATGCTCTTTTGAGAAATTTGAAAGGGT 5002
884 uMetLeuProTrrpGlyCysAspValGlnLysLeuSerSerLeuLeu 901
5003 TGTTCAGCAAAACAGTGCAGTGAATGAAGTATTTTGCACAGAGCTGC 5052
901 yHis.....LeuGluGluValProGlnLeuValLysLeuGly 913
5053 AGGAGAGAAATCTGACCAAGAAACCCCTTGGCCCCCAGCAAGGGT 5102

```

```

914 LeuLysAsnTrrpArgLeuThrsAspThrGluIleArgIleLeuGlyAlaPhe 930
5103 GTTGA.....GAGCTTGAGTCCCTCTGTGAGCT 5131
930 ePheGlyLysAsnProLeuLysAsnPheGlnGlnLeuAsnLeuAlaGlyA 947
5132 GTACCG.....CAGTCCCTCGAAGAGCTGAGAG 5160
947 snArgValSerSerAspGlyTrrpLeuAlaPheMetGlyValPheGluAsn 963
5161 AGAAATGCAAACTCAGGAGTATGAAAAATAGGAATT...CAAGAG 5207
964 LeuLysGlnLeuValPhe.....PheAspPheSerThrLys 975
5208 CTCGACAGCTTATTAAGTTCTGAAGCAGAGAGCTTGCTGCTTAGACA 5257
975 sGluPheLeuProAspProAlaLeuValArgLysLeuSerGlnValLeuS 992
5258 GCAGTATTTGCAGAAAT.....GACAGTGGCAGACGAAAGCTGA 5298
992 eLysLeuThrsPheLeuGlnGluAlaArgLeuValGlyTrrpGlnPheAsp 1008
5299 CAAGCGTGACTCTGGAGATGAGTCCAAAGTTGCGGCGAGAAAAAGAAACAG 5348
1009 AspAspAspLeuSerVal 1014
5349 ACGGAAACACTCTCCTT 5366

seq_name:/cgn2_6/plodata/2/lna/6A_COMB.seq:us-08-947-823-1
seq_documentation block:
: Sequence 1, Application US/08947823
: Patent No. 6114605
: GENERAL INFORMATION:
: APPLICANT: Williamson, Valerie M.
: APPLICANT: Kaloshian, Isqouhi
: APPLICANT: Yaghoobi, Jafar
: APPLICANT: Bodeau, John
: APPLICANT: Milligan, Stephen
: TITLE OF INVENTION: Procedures and Materials for Conferring
: NUMBER OF SEQUENCES: 5
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend and Crew LLP
: STREET: Two Embarcadero Center, Eighth Floor
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94111-3834
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/947, 823
: FILING DATE: 09-Oct-1997
: CLASSIFICATION: 800
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/US97/18802
: FILING DATE: 09-Oct-1997
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/028, 191
: FILING DATE: 10-Oct-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Bastian, Kevin L.
: REGISTRATION NUMBER: 34,774
: REFERENCE/DOCKET NUMBER: 023070-070210US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 576-0200
: TELEFAX: (415) 576-0300
: INFORMATION FOR SEQ ID NO: 1:

```



```

415 rValaAsnGluAspValLeuLeuThrThrGlyLeuLeuCysLysTyrThrA 432
      ::::: |||||
1980 TACTACTACATGGAGTACGTTAACA..... 2004
432 lacGlnArgPheLysProLysTyrLysPhePheHisLysSerPheGlnGlu 448
      :::: ||| |||
2005 .....AGACCTTTCTCTGAA 2019
449 TyrThrAlaGlyArgArgLeuSerSerLeuLeuThrSerHisGluProGlu 465
      ||| ||||| :::: ||| |||
2020 TCTAAGACGAAGAGTAGAGAT.....ATTGTGACAACTCGGGAAGAAAGA 2063
465 uGluValThrLysGlyAsnGlyTyrLeuGlnLysMet...ValSerIles 481
      | :::: ||||| ::::
2064 AGTGGCTTTGCATGGAAAGCTGAACTGACACTGATCCTTGCCTTCGATTC 2113
481 erAspIleThrSerThrTyrSerSerLeuLeuArgTyrThrCysGlySer 497
      :::: ||||| :::: |||||
2114 TAAGACCAATGAAAGTTGGAACTATTAGAGAAAGGCGCATTTGGGAT 2163
498 Ser.....ValGluAlaThrArgAlaValMetLysHis 508
      :::: ||||| :::: |||||
2164 GAGAGTTGCCCTGATGAACTATTAGATGTCGGTAAAGAAATAGCGGAAA 2213
508 s.....LeuAlaAlaValTyrGlnHisGlyCysLeuLeuG 520
      : ||||| |||||
2214 TTGTAAAGGCTTCCTTGTGGTGTGCTGATGATGCTGGAGTCATCTGCTG 2263
520 lLeuSerIleAlaLysArgProLeuTyrParg.....GlnGluSerLeu 534
      || :::: ||||| :::: |||||
2264 GGAGGGAA...AAGAAAGAGAGTGTGTGCTGAGTCAAGTCAAGTGTGG 2310
535 GlnSerValLysAsnThrThrGlnGlnGluLeuLysAlaIleAsnIle 551
      :::: ||||| :::: |||||
2311 AATTCTTTTATTGTAACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 2360
551 eAsnSerPheValGlyCysGlyIleHisLeuTyrGlnGluSerThrIle 568
      ::::
2361 AAGT..... 2364
568 ySerAlaLeuSerGlnGluPheGlnAlaPhePheGlnGlyLysSerLeu 584
2364 ..... 2364
585 TyrIleAsnSerGlyAsnIleProAspTyrLeuPheAspPheGlnIle 601
      :::: |||
2365 .....TATGACCA 2372
601 sLeuProAsn.....CysAla..... 606
      |||||
2373 TTATCCATCACTACCTCAAGCCATGCTGCTGATTTTGCAGATTTTCCGA 2422
607 .....SerAlaLeuAspPheIleLysLeuAspPheTyrGlyAlaMet 621
      :::: ||||| :::: |||||
2423 AGGACACTTCATTTGACAACTCATGAGTGAATGATTATTTGGTGTGGA 2472
622 AlaSerTyrGluLysAlaIleGluAspThrGlyGlyIleHisMetGlu 638
      :::: ||||| :::: |||||
2473 GGATTTGTGGGAAGACGAGATGAACACT.....ATGGAAGA 2510
638 uAlaProGluThrTyrIleProSerArgAlaValSer.....LeuP 652
      ||:: ||||| :::: |||||
2511 AGTGGGAGAGATTTATATGATGATTTAATTACAGTACCTGGTAATTT 2560
652 hPheAsn.....TyrLysGlnGluPheArgThrLeuGluValThr 665
      ||||| :::: |||||
2561 GTTTCAATGAGATAGGTATGACATGAATTTCCAATTCATGATCTT... 2607
666 LeuArgAspPheSerLysLeuAsnLysGlnAspIleThrTyrLeuGlyLys 682
      :::: ||||| :::: |||||
2608 GTGCATGACTTTTGTGATTAAGCAAGAAAGAAATTTGTTGATCA 2657

```

```

682 sIlePheSerSerAlaThrSer.....LeuArgLeuGlnIle..... 694
      ::||| ||||| ||| |||
2658 GATAGATCAAGTCTCCATCAGATTTGTGCTGCTGCAAAATTTACCATTC 2707
694 ..... 694
2708 ATTGTGATGAGGAGGACACTTTGGCTTAATTTGTCATGTTGCATTTCA 2757
695 .....LysArgCysAlaGlyValAlaGlySerLeuSerLeuValLeuSe 709
      ||||| :::: |||
2758 AATAAGAAAGGCATCTCGT..... 2778
709 rThrCysLysAsnIleTyrSerLeuMetValGluAlaSerProLeu... 724
      ||||| :::: |||||
2779 .....AAACCTCTATTCTTTGAGCATATTGGACACAGCTGGATG 2821
725 ..ThrIleGluAspGluArgHisIleThrSerValThrAsnLeuLysThr 740
      :::: ||||| :::: |||||
2822 ACAGTGTTCGTGATGCATTTGCCTAAGACACTTGAGGCTCTTATGAGTG 2871
741 LeuSerIleHisAspLeuGlnAsnGlnArgLeuProGlyGlyLeuThrAs 757
      ||||| :::: |||
2872 TTGGACCTGCAT..... 2883
757 pSerLeuGlyAsnLeuLysAsnLeuThrLysLeuIleMetAspAsnIle 774
      ||:: |||||
2884 .....ACGCTCTTTATTCATG..... 2898
774 yMetAsnGlnGluAspAlaIleLysLeuAlaGluGlyLeuLysAsnLeu 790
      ::||| :::: |||
2899 .....GTGAAGATCTTGG.....CTG 2916
791 LysLysMetCysLeuPhe...HisLeuThrHisLeuSerAspIleGly 806
      :::: ||||| :::: |||||
2917 AATGAATAATGATGTGATCAATCAATGAGTCAATATCC..... 2955
806 uGlyMetAspTyrIleValLysSerLeuSerSerCysLeuProCysAspLeu 823
      ::||| |||||
2956 .....ATTGACACACAAGTTAAATATCTG.....CCTTGTGCTTTCT 2992
823 lGluIle...GlnLeuValSerCysCysLeuSerAlaAsn...AlaVal 837
      :::: ||||| :::: |||||
2993 CAACCTCTGGAATGTAGAAAGCCTGTTGTGTACCAACAGATCAATC 3042
838 LysIleLeuAlaGlnAsnLeuHisAsnLeuValLysLeuSerIleLeuAs 854
      ::||| :::: |||||
3043 TTGGTACTATTAACCGAATTTTGGATCTGTGAAGTTCGACGATCTGTGTC 3092
854 pLeuSerGlu...AsnTyrLeuGluLysAspGlyAsnGlnAla..... 867
      :::: |||||
3093 CGTGGATGCTGTGTTCTTTGATGATGATGACATGAAATCAATATTTGA 3142
868 .....LeuHisGluLeuIle 872
      ||| |||||
3143 TAGCAGAGGACACAAAGTTAGAGACTTGAGATTTTAAACGAACTGTG 3192
873 .....AspArgMetAsnValLeuGlnGluLeuThrAlaLe 884
      ||| :::: |||
3193 ATTTCCTATTGGAAGATCAAAAGAAATTTTCAAAAGTTTCCCAATCT 3242
884 uMetLeu.....ProTyrGlyAspValGln 894
      ||:: |||||
3243 TCAGTTGCTTCATTTGAACTCAAGAGACTCATGGATTTATTTCAACAGAGC 3292
894 lYserLeuSerSerLeuLeuLysHisLeuGlnGluValProGlnLeuVal 910
      ||| |||||
3293 AACATGTGTTCTCGGAATTTGGATTTCCCTAAGTGAAGTCA...GAAACAGCTC 3339
911 LysLeuGlyLeuLysAsnTyrPargLeuThrAsp..... 921
      :::: ||||| :::: |||||
3340 TCTGTAGGTTTAAAGTTCAACACAAACAGATAGTGGTCTCTGTAGAC 3389
922 .....ThrGluIleArgIleLeuG 928

```

```

3390 GACAAATCGCCGTGGATTTCCTCCCTCAAAATTTGAAATAGTCTG. 3438
928 LylalaphhegilyLysasnPro..... 925
3439 .....TCGTCGCTGAATTTCCCGCTGCATCCGATTCACATATCAACATA 3483
936 .....LeulysasnphengingInleuAsnleuAlaGlyAsnArgAlaSe 950
3484 GCGAGACTGCCCACTTGAAGAGTTGTCCCTTATCATCAATCAATCA 3533
950 rSerAspGlyTTP...LeulalaphemetGlyValPhegluAsnleuysG 966
3534 TGGAGAGAAATGGACATGGGAGGAGAAACACCTTTGAGAATCTCAAT 3583
966 InleuValPhepaspPheSerThrIysgluphe 977
3584 TTTTGAACCTTCATCAAGTTAGTATTTCCAAGTGG 3618

```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-180-439-5

documentation_block:

Sequence 5, Application US/09180439

Patent No. 6225532

GENERAL INFORMATION:

APPLICANT: Dixon, Mark S

APPLICANT: Hatzixanthis, Kostas

APPLICANT: Jones, David A

APPLICANT: Jones, Jonathan DG

TITLE OF INVENTION: Plant pathogen resistance genes and uses thereof

FILE REFERENCE: 620 - 53

CURRENT APPLICATION NUMBER: US/09/180,439

EARLIER FILING DATE: 1998-12-06

EARLIER APPLICATION NUMBER: PCT/GB97/01249

EARLIER FILING DATE: 1997-05-08

EARLIER APPLICATION NUMBER: GB 9609681.3

EARLIER FILING DATE: 1996-05-09

EARLIER APPLICATION NUMBER: GB 9619924.5

NUMBER OF SEQ ID NOS: 10

SOFTWARE: Patent In Ver. 2.0

SEQ ID NO 5

LENGTH: 3541

TYPE: DNA

ORGANISM: Lycopersicon esculentum

US-09-180-439-5

alignment_scores:

Quality: 143.50

Ratio: 0.755

Percent Similarity: 50.132

Percent Identity: 23.747

US-09-697-089-2 x US-09-180-439-5 ..

Align seg 1/1 to: US-09-180-439-5 from: 1 to: 3541

```

675 GlnAspIleThrIleuGlyLysIlePheSerSerIleThrSerLeuAr 691
1227 GAAGAAATAGTACTTA.....AGTCTCTTACTTAAGCTATC 1264
691 GLeuGlnIleLysArgCysAlaGlyAlaGlySerLeuSerLeuVal 708
1265 TTGGATATCAAC.....TTTCTTAGTGTTCATTCCTGCTTCAT 1305
708 euserThrCysAsnIleTyrSerIleuMetValGluAlaSerProLeu 724
1306 TGGGGAATTCGAACAACCTGCTTTTGTATCTTACAAATAATCAAGCTT 1355
725 .....ThrIleGluAspGluArgHisIleThrSerValThrAsnLeu 739
1356 TCTGGCTCTATCTCTGAAGAA.....ATAGCTTACTCTAAG 1390

```

```

739 sThrIleSerIleHisAspLeuGlnAsnGlnArgLeuProGlyGlyLeu 756
1391 GTCTCTTACTTACCTAGATTGTAAGAGAAATGCTCTTAATGGCTATATC 1440
756 hArgSerLeuGlyAsnLeuLysAsnLeuThrIleLysLeuIleMetAsp 772
1441 CTGCTTCATTTGGGGAATCTGAACAACCTTGCTAGGTGATCTTACAA 1490
773 IleLysMetAsnGluGluAspAlaIleLysLeuAlaGlyLeuLysAs 789
1491 AATCAGCTTCTTGGC.....TCTATTCTGGAAGAAATAGCTTA 1528
789 nLeuLysLysMetCysLeuPheHisLeuThrHisLeuSerAspIleGly 805
1529 CTGAGACTCTCTACTAATCTATATTTGGGTATTAATCTCTTATTTGAC 1578
806 .....Glu 806
1579 TTAATCTCTGCTTCATTCGGAATATGAGAAATCTGCAAGCTGTTTCTC 1628
807 GlyMetAspTyrIleValLysSerLeuSerSerGluProCysAspLeuG 823
1629 AATGATTAACAATCTCATTTGGGAAATTCCTTCATTTGTGTGCATTTAAC 1678
823 uGluIleGlnLeu.....ValS 829
1679 ATCATAAGAACTGTGTATATGCGGAGAAACAATTGAAGGCAAAAGTTC 1728
829 erCysCysLeu.....SerAla 834
1729 CGCAATGTTGGGTAATATACAGTACGCTTCGTTGTTGCATGTCATCT 1778
835 AsnAlaValLys...IleLeuAlaGlnAsnLeuHisAsnLeuValLysLe 850
1779 AATAGTTTCACTGAGAGAGCTCCCTTCATCTATTTCCAAATTTAACATC 1828
850 uSerIleLeuAspLeuSerGluAsnTyrLeuGluLysAsp..... 863
1829 AAAAATACTGATTTTGGCGAACAATCTCGAGGAGCAATACCACAAT 1878
864 ....GlyAsnGluAlaLeuHisGluLeuLeuAspArgMetAsnValLeu 878
1879 GTTTTGGCAATATATATAGTCCAGCTTGTGATGATGAGCAAT..... 1922
879 GluGlnLeuThrAlaLeuMetLeuProTyrGlyCysAspValGlnLys 895
1923 AACAACTTCTGGG...ACTCTCCACAACAATTTAGCATTTGGATGTTTC 1969
895 rLeuSerSerLeuLeuLysHis..... 902
1970 ACTGATTAAGTCTCACTTGCATGCATGCAATGAATAGAGATGAATCCCT 2019
903 ....LeuGlnGluValProGlnLeuValLysLeuLysLeuAsnTyr 917
2020 GTCTTTTGGACAATTCGAAAAAGCTGCAAGTCTGTGATTTAGGAGACA 2069
918 ArgLeuThrAspThr.....GluIleAr 925
2070 CAACCTCAAGCACACATTTCCCATGCTGTTGGAGACTTTGCCAGAGCTG 2119
925 gIleLeuGlyAlaPhePheGlyLys.....AsnProLeuLys..... 937
2120 AGTTTAAAGTTGACATCGAATAAATTGCATGAGCACTATAGATCATCAG 2169
938 .....AsnPheGlnGlnLeuAsnLeuAlaGlyAsn 947
2170 GGGCTGAATATCATGTTCCATGATCTTGATCATATGATCTCTCTCCCA 2219
948 ArgValSerSerAspGlyTyrLeuAlaPheMetGlyValPheGluAsn 964
2220 GCATTCTCGCAAGAC.....TTACCAACAAGCTATTTGAAACATTT 2260

```


1830 GATCAAGATATGACGACAAAGATTGTTGATACAAATTTTCAGTCAAGT 1879
 536 erval 537
 1880 TAGTGGCTCATGATCAATTTGAGTGAATATTGATGTTGCTGATAAT 1929
 538 LysAsnThrThrGluGluIleLeuLysAlaIleAsnIleAs 552
 1930 TGGGAAACAAACGTGTTGAAAGAGATCTTAATG..... 1965
 552 nSerPheValGluCysGlyIleHisLeuTyr..GlnGluSerThr.... 566
 1966 .TCTTATGATGATGTGTGGTACTACTACATTTGGATGATGGACAAAGACC 2014
 567SerLysSerAlaLeu...SerGlnGluP 575
 2015 TTTTCCGAGCTAAGAAAGAACTAGATTTATTTGACAACTGAGAA 2064
 575 heGluAlaPhePheGlnGlyLysSerLeuTyrIleAsnSer..... 588
 2065 AGGAAAGTGGCTTTCATGGAAG.....CTGACACTGATCTCTT 2105
 588 588
 2106 GACCTTCGATTGCTAAGACAGATGAAAGTTGGACCTTTAGATAAAG 2155
 589GlyAsn.....IleProAspTyrLeuPheAspPheGluH 601
 2156 GACATTGTGATATGAGAGTGGCTGATGAACTATTGATGTGCGTAAAG 2205
 601 ILeuProAsnCysAlaSerAlaLeuAspPheIleLysLeuAspPheTyr 617
 2206 AAATAGCCGAAATTTGAAAGGCTCTCTTGGTG...GCTGATCTGATT 2252
 618 GlyGlyAlaMetAlaSerTyrGluLysAlaIleGluAspThrGlyIle 634
 2253 GCTGGAGTCAATGCTGGAGGAGAAAG.....AAAGAGTGTGTGCT 2296
 634 eHsMeGluGluAlaProGluThrTyrIleProSerArgAlaValSerL 651
 2297 TGAAGTCAAGTAGTTGAGTCTTTATTTTGACAGTCAAGTGGAG 2346
 651 eu..... 651
 2347 TGATGAAAGTTAATAGATTATGACATTACCATCACCCTCAAG 2396
 652PhePheAsnTrp..... 655
 2397 CCATGCTTCTTCACTTTCGCAAGTTGGCCGAGACACTCCTTGACAA 2446
 656LysGlnG 658
 2447 CTATTTTACTCTTTATTGGTCTGAAAGATTGTGAAAGAACGG 2496
 658 IuPheArgThrLeuGluValThrLeuArgAspPheSerLysLeuAsnLys 674
 2497 AGATGAGGCTAATAGAGAGAGTGTGAAG..... 2525
 675 GlnAspIleThrTyrLeuGlyLysIlePheSerSerAlaThrSerLeuAr 691
 2526ATTATATGATGATTATTTCCAGTAGCTTGTAAT.. 2564
 691 glEuGlnIleLysArgCysAlaGlyValAlaGlySerLeu..... 704
 2565TGTTCATGAGATAGTGATATACTGAAATTTCC 2598
 705SerLeuValLeuSerThrCys.....Lys 712
 2599 AAATTCATGATCTGTGCATGACCTTTGTTGATAAAGCAAGAAAGAA 2648
 713 AsnIleTyrSerLeuMetValGluAlaSerPro..... 723
 2649 AATTTGTTGATCGATAGATCAAGTCTCCATCAGATTTGTTGCCCTCG 2698

724LeuThrIle.....GluAspGluArgHis..... 731
 2699 TCAATATTCATCATCATATATGATGAGGAGAGACACTTGTGGCTTAAT 2748
 732 ..IleThrSerValThrAsnLeuLysThrLeuSerIleHisAspLeuGln 747
 2749 TTGTCAATGTTGCAATTCAAATAGAAAGGCATCTGTGTAACACCTTAT 2798
 748 AsnGlnArgLeuProGlyGly..LeuThrAspSerLeuGlnLysLeu 763
 2799 TCTTTGAGATAAATGAGACACCTGATGACAGTCTTCTGATGCATT 2848
 763 AsnLeuThrLysLeuIleMetAspAsnIleLysMeCAsnGluLysAs 780
 2849 TCACCTAAGACACTTGGAGCTTATTAGAGTGTGACCTGGAACCCCTT 2898
 780 IalIleLysLeuAlaGluGlyLeuLysAsnLeuLysLysMetCysLeuPhe 796
 2899 TAATCATGCTGAATGATCTTTG.....CTGAATGAAATATGCAATGTTG 2942
 797 ..HisLeuThrHisLeuSerAspIleGlyGluLysMetAspTyrIleVa 812
 2943 AATCATTTGAGTACTTAAAG...ATTGCGACACACCTTAATATCTGCC 2989
 812 LysSerLeuSerSerGluProCysAspLeuGluIleGlnLeuValS 829
 2990 TTTTCTCTTCCAAAC...CTCTGGAATCTAGAAAGTCTG...TTGTGT 3033
 829 erCysCysLeuSerAlaAsnAlaValLysIleLeuAlaGlnAsnLeuHis 845
 3034 CT.....AACAAAGATCAATCTTGGTACTATTACCGAGAAATTTTG 3074
 846 AsnLeuValLysLeuSerIleLeuAspLeuSerGlu...AsnTyrLeuG 861
 3075 GATCTGTAAAGTTGCGAGTGTCTCCGTGGTCTTCTTCTTTTGA 3124
 861 uLysAspGlyAsnGluAlaLeuHisGluLeuIleAspArgMetAsnVal 878
 3125 TATGATGCGAGATCAATCAATA.....TTGATATGCAAGAGACCAAGT 3168
 878 euGlnGluLeuThr.....AlaLeuMetLeuProTyrPylCysAsp 891
 3169 TAGAGAACTTGAATATTAGGGGACCTGTGATTCTTATTCGAAAGAT 3218
 892 ValGlnGlySerLeuSerSerLeuLeuLysHisLeuGluGluValProG 908
 3219 ACAATG.....AATATTTCAAAAGGTTTCCCAATCTT..CA 3253
 908 nLeuValLysLeuGlyLeuLys..AsnTyrParg..... 918
 3254 GGTGCTTCAGTTGACCTCAAGAGCATGAGATTATTCAACAGAGCAAC 3303
 919LeuThrAspThrGluIleArgIleLeu 927
 3304 ATTGCTTCCGGAATGGAATGCTTAACTGACACTGAACCACTCTGTGA 3353
 928 GlyAlaPhePheGlyLysAsnProLeuLysAsnPheGlnGluLeuAsnLe 944
 3354 GGT.....TTTAAAGTTCAAACACCAACCAACCTG 3382
 944 uAlaGlyAsnArgValSerSerAspGlyTyrPheAlaPheMetGlyValP 961
 3383 TGGCTCTCTGTGTTGACAAATGCGCGTGGGATTTCACCTCCCT.... 3428
 961 heGluAsnLeuLysGlnLeuValPhePheAspPheSer...ThrLysGlu 976
 3429 ..TCAATTTGAAAGAACTGTTGTTATGACTTCTCTGACATCCGAT 3476
 977 PheLeuProAspProAlaLeuValArgLysLeuSerGlnValLeuSerLy 993
 3477 TCACATATCAACAAATAGAGAGACTGCCCAACCTTGAAATTTGTCCTTTA 3526


```

1122 GAAACTT.....CACCTTGCTGCTGCT 1147
756 hrAspSerLeuGlyAsnLeuLysAsnLeuThrLysLeuIleMetAspAsn 772
      :|||||:|||||:|||||:|||||:
1148 TGGAGAAACCTGAAACCTTCAGACACTTGAT.....TTAAGCCAT 1188
773 IleLysMetAsnGlnGluAspAlaIleLysLeuAlaGlnGlyLeuLysAs 789
      :|||||:|||||:|||||:|||||:
1189 AATGACATTAAGAGGCTCTGACTGCTGACGCTGCATA.....CTCAAAA 1232
789 nLeuLysLysMetCysLeuPheHisLeuThrHisLeuSerAspIleGly 806
      :|||||:|||||:|||||:|||||:
1233 CCTGTCCCACTTGCAAACTTTAACTGAGCCCAATAGAGCCTTGCT 1281
806 LuGlyMetAspTyrIleValLysSerLeuSerSerGluProCys...Asp 821
1282 .....CTCCAGAGTCAGGCAATTCAAAGATGCTCCAG 1314
822 LeuGlnGlnIleGlnLeuValSerCysCysLeuSerAlaAsnAlaVally 838
      :|||||:|||||:|||||:|||||:
1315 CTAAACTCTCGATTGGCATTTACCCGCTTACACATTAATGCTCCACA 1364
838 sIleLeuAlaGlnAsnLeuHisAsnLeuValLysLeuSerIle..... 852
1365 AAGTCCCTTCAAAACCTCCATTTCTTCAGGTTTGAATCTCACTACT 1414
853 .....LeuAspLeuSerGlu..... 857
1415 GCTTCCTTGATACAGCAATCAGCATCTTCTAGCAGGCTACAGTTCTC 1464
858 AsnTyrLeuGlnLysAspGlyAsnGlnAlaLeuHisGlnLeuIleAsp 874
      :|||||:|||||:|||||:|||||:
1465 CGGATCTCTCAACTTAAAGGAATCACTTTCAGATGGAGTATACAGAA 1514
874 gMetAsnValLeuGlnLeuThrAlaLeu.....MetLeuProIle 889
      :|||||:|||||:|||||:|||||:
1515 GACCAACCTTCTAGACCGTGGCAGCTTGAGGCTTGATTTGCTCT 1564
889 LysCysAspValGlnGlySerLeuSerSerLeuLeuLysHisLeuGln 905
      :|||||:|||||:|||||:|||||:
1565 CTCTCT.....GGTCTCTCTCTATAGACCAAGCATTCACAGC 1605
906 ValProGlnLeuValLysLeuGlyLeuLysAsnTyrParGlyLeuThr 922
      :|||||:|||||:|||||:|||||:
1606 TTGGGAAAATGAGCCATGTAGACTTAAGCCACACAGCCGATCATGCGA 1655
922 gLulIleArgIleLeuGlyAlaPhePheGlyLysAsnProLeuLysAsn 939
      :|||||:|||||:|||||:|||||:
1656 CAGCATTGATCTCTAGCCATCTTAAGGA..... 1686
939 heGlnGlnLeuAsnLeuAlaGlyAsnArgValSer 950
1687 ..ATCTACCTCAATCTGCTGCCAAGACAGATTAAAC 1719
seq_name: /cgn2.6/ptodata/2/lna/5B_COMB.seq:US-08-833-823-3
seq_documentation_block:
; Sequence 3, Application US/08833823
; Patent No. 5969093
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John
; APPLICANT: Kelleher, Kerry
; APPLICANT: Carlin, McKeough
; TITLE OF INVENTION: DNA SEQUENCES AND SECRETED PROTEINS
; TITLE OF INVENTION: ENCODED THEREBY
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc. -- Legal Affairs
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: USA

```

```

; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/833,823
; FILING DATE: 10-APR-1997
; CLASSIFICATION: 530
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 08/514,014
; FILING DATE: 11-AUG-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; REFERENCE/DOCKET NUMBER: G16000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2582 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 52..2034
; US-08-833-823-3

```

```

alignment_scores:
  Quality: 142.00      Length: 545
  Ratio: 0.553        Gaps: 24
  Percent Similarity: 47.156  Percent Identity: 22.936
alignment_block:
US-09-697-089-2 x US-08-833-823-3 ..

```

```

Align seg 1/1 to: US-08-833-823-3 from: 1 to: 2582
486 ThrTyrSerSerLeuLeuArgTyrThrCysGlySerSerValGlnAla 502
      :|||||:|||||:|||||:|||||:
265 ACCTTCAGACAGACTCATGATCTTACC.....TTTGGATTAAAC 305
502 rArgAlaValMetLys.....HisLeuAlaAlaValTyrGlnHisGly 517
      :|||||:|||||:|||||:|||||:
306 TAGGTGCCAGATTAACTGATCATGAAGACACTTTCAAAAGCCATCATC 355
517 LysLeuLeuGlyLeuSerIleAlaLysArgProLeu.....TyrArgGln 531
      :|||||:|||||:|||||:|||||:
356 AATTAAAGCACACTTGTTACTGGAATCCCTGATATTTCATGCGAGAA 405
532 GluSerLeuGlnSerValLysAsnThrThrGlnGlnGlnLeuLysAl 548
      :|||||:|||||:|||||:|||||:
406 ACATCGCTTAATGAGCCCAAGTCA.....CTGAGG.. 435
548 aIleAsnIleAsnSerPheValGluCysGlyIleHisLeuTyrGlnLys 565
      :|||||:|||||:|||||:|||||:
436 .....CANCCTTTCTTAATCC 451
565 eThrSerLysSerAlaLeuSerGlnPhe.....GluAlaPhe 578
      :|||||:|||||:|||||:|||||:
452 AAACGGGAATATCCAAATCTC.....GAGTTATTCCAGTGCACAAATCG 495
579 PheGlnGlyLysSerLeuTyrIleAsnSerGlyAsnIleProAspTyrLe 595
      :|||||:|||||:|||||:|||||:
496 GAAACCTTGAAACCTTGATCTGTAAGAGCAGACCATATTTCCTCATTA 545
595 uPheAspPhePheGlnHisLeuProAsnCysAlaSerAlaLeuAspPhe 612

```



```

546 GTTC.....CCCAAGACTTCCAGCAGCAATCTGA 577
612 IeLysLeuAspHeterglyglyAla..... 620
578 AAGTACTGGATTTCAGAAATAGCTATACACTCTCTAGAGAAAGC 627
621 MetAlaSerTrpGluLysAla.....Ala1 629
628 ATGAGGTCTCTGGAGCAGGCCATCAACCTAAGCTGAATTCAGAGCA 677
629 uAspHeterglyglyLeuHisMetGluGluAlaProGluThrTyrIlePro 646
678 TAATGTTAAAGSTATTGAGCTTGGGCTTTTGATTCAACGGTCTCCAA 727
646 er.....ArgAlaValSerLeuPheAsnTrpLys 656
728 GTTGAACATTGGAGCACTCCAAATTTGCTGTATATATCAATGGCTCG 777
657 GluGluPheArgThrLeuGluValThrLeuArgAspPheSerLysLeu 673
778 CAGAACTCTACTACAGTCTCTGGCTGGGAGACATTGGAGACATTGA 827
673 nLysGlnAspIleThr.....TyrLeuGlyLysIlePheSerSer 687
828 TGACGAAGATTATTAGTTCAGCCATGCTCAAGGACTCTGAATGTCTG 877
687 IaThrSerLeuArgLeuGlnIleLysArgGlyAlaIaGlyLys 703
878 TTGAGAGCTCAACCTGCAGGAACACGCTTCTGACATCTCATCCAC 927
704 Leu..... 704
928 ACATTTCAGTGTCCACCACCTCCAGAAATTGATCTGACAGCACTCA 977
705 .....SerL 706
978 CTTGAAGGGTTACCCCTCTGGGATGAGGGTCTGAACCTGCTCAAGAAAT 1027
706 euValLeuSerThrCysLysAsnIleTyrSerLeuMetValGluAlaSer 722
1028 TAGTTCCTCAGTGA.....AATCATTTGATCAATTTGTGTCAAAATCACT 1071
723 ProLeuThrIleGluAspGluArgHisIleThrSerValThrAsnLeu 739
1072 GGTGCCAATTTCCCTCCCTTACACACCTTACATCAGAGGCAACGTGA 1121
739 sThrLeuSerIleHisAspLeuGlnAsnGlnArgLeuProGlyGlyLeu 756
1122 GAACCT.....CACCTTGGCTTGGCTGCT 1147
756 hrAspSerLeuGlyAsnLeuLysAsnLeuThrLysLeuIleMetAspAsn 772
1148 TGGAGAAACTAGAAACCTTCAGACACTTGAT.....TTAAGCCAT 1188
773 IleLysMetAsnGluGluAspAlaIleLysLeuAlaGluGlyLeuLysAs 789
1189 AATGACATAGAGGCTTCTAGCTGCTGCA.....CTCAAAA 1232
789 nLeuLysLysMetCysLeuPheHisLeuThrHisLeuSerAspIleGly 806
1233 CCGTGCCCACTTGCAAACTTAAACCTGAGCCCAATGAGCCTTGCT 1281
806 LuGlyMetAspTyrIleValLysSerLeuSerGluProCys...Asp 821
1282 .....CTCCAGAGTCAAGGCATTCAGAAAGATGCTCCTCAG 1314
822 LeuGluGluIleGluValSerCysLeuSerAlaAsnAlaVal 838
1315 CTAGAACTCCTGATTTGGCATTTACCCGCTTACACATTAAATGCTCCACA 1364
838 sIleLeuAlaGlnAsnLeuHisAsnLeuValLysLeuSerIle..... 852

```

```

1365 AAGTCCCTTCCAAACCTTCATTTCCCTTCAGGTTCTGAATCTCATTACT 1414
853 .....LeuAspLeuSerGlu..... 857
1415 GCTTCCTTGATACAGCAATTCAGCATCTTCTAGCAGGCTTACCAGTTCTC 1464
858 AsnTyrLeuGluLysAspLysnGluAlaLeuHisGluLeuIleAspAr 874
1465 CCGCATCTCACTTAAAGGAATCACTTTCAGAGATGGGACTATCAGAA 1514
874 gMetAsnValLeuGluGluLeuThrAlaLeu.....MetLeuProTrp 889
1515 GACCAACCTACTTCAGACCGGTGGAGCTTGGAGGTTCTGATTTGTCTCT 1564
889 LysAspValGlnGlySerLeuSerSerLeuLeuLysHisLeuGlu 905
1565 CTGTG.....GCTCTCTCTATAGACCCAGCAACGATTCACACAC 1605
906 ValProGluLeuValLysLeuGlyLeuLysAsnTrpArgLeuThrAsp 922
1606 TTGGGAAATAGGCCATGTACACTTAAGCCACACAGCTGACATGCA 1655
922 rGluIleArgIleLeuGlyAlaPhePheGlyLysAsnProLeuLysAsn 939
1656 CAGCATGTATCTCTTACGCATCTTAAGGA..... 1686
939 heGlnLeuAsnLeuAlaGlyAsnArgValSer 950
1687 ..ATCTACCTCAATCTGGCTGCCAACAGCATTAAC 1719

```

THIS PAGE BLANK (USPTO)

CC Ischemia/reperfusion injury, excitotoxic brain damage, acute bacterial
 CC meningitis and liver disease.
 XX
 SQ Sequence 3133 BP; 903 A; 691 C; 729 G; 810 T; 0 other;

alignment_scores:
 Quality: 5281.00 Length: 1024
 Ratio: 5.157 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-697-089-2 x AAS03945 ..

Sign seg 1/1 to: AAS03945 from: 1 to: 3133

1 MetAspPheIleLysAspAsnSerArgAlaLeuIleGlnArgMetGlyMe 17
 |||||||
 36 ATGATTTTCATTAAGCAATACGACAGTCCCTTATTCATAAGAAATGGCAT 85
 17 tThrValIleLysGlnIleThrAspAspLeuPheValTrpAsnValLeu 34
 |||||||
 86 GACGTGTAAAGCAATACACAGATGACCTATTGTATGAAATGTTCTGA 135
 34 snArgGluGluValAsnIleIleCysGluLysValGluGlnAspAla 50
 |||||||
 136 ATCCGGAAGAAATTAACATTTTCTCTCGAAGAGTGAGACAGATGCT 185
 51 AlaArgGlyIleIleHisMetIleLeuLysGlySerGluSerCysAs 67
 |||||||
 186 GCTAGAGGATCATTCACATGATTTTGGAAAAAGGTTTCAGACTCCTGTAA 235
 67 nLeuPheLeuLysSerLeuLysGluTrpAsnTrpProLeuPheGlnAsp 84
 |||||||
 236 CCTCTTCTTAATCCCTTAAGAGATGGAACTATCCTTATTTCCAGACT 285
 84 euAsnGlyGlnSerLeuPheHisGlnTrpSerGluGlyAspLeuAsp 100
 |||||||
 286 TGAATGGCAAAAGTCTTTTTCATCAGACATCAGAAGGAGACTTGGACGAT 335
 101 LeuAlaGlnAspLeuLysAspLeuTrpHisTrpProSerPheLeuAsnPh 117
 |||||||
 336 TTGGCTCAGATTTAAAGGACTTGTACCATACCCATCTTTTCTGAACCT 385
 117 eTrpProLeuGlyGluAspIleAspIleIlePheAsnLeuLysSerTrp 134
 |||||||
 386 TTATCCCTTGGTGAAGATATGACATTATTTTAACTTGAAGAACACCT 435
 134 heThrGluProValLeuTrpArgLysAspGlnHisHisArgValGlu 150
 |||||||
 436 TCACGAACCTGTCTCTGGAGGAAGACCAACACCATCACCGCTGGAG 485
 151 GlnLeuThrLeuAsnGlyLeuLeuGlnAlaLeuGlnSerProCysAla 167
 |||||||
 486 CAGCTGACCTGAATGGCTCTCTGCAAGCTCTTCAGAGCCCTCGATCAT 535
 167 eGluGlyLysSerGlyLysGlyLysSerThrLeuLeuGlnArgIleAla 184
 |||||||
 536 TGAAGGGGAATCTGGCAAGGCAAGTCCACTCTGCTGCAGCGCATTTGCCA 585
 184 eLeuTrpGlySerGlyLysCysLysAlaLeuThrPheLysPheAla 200
 |||||||
 586 TGGCTGTGGGGTCCGGAAAGTGAAGGCTGTGACCAAGTTCAAATTCCTGC 635
 201 PhePheLeuArgLeuSerArgAlaGlnGlyLeuPheGluTrpLeuCys 217
 |||||||
 636 TTCTTCTCTCCGCTCAGCAGGCGCCAGGCTGACCTTTTGAACCCCTCTG 685
 217 sAspGlnLeuLeuAspIleProGlyThrIleArgLysGlnTrpPheMet 234
 |||||||
 686 TGATCAACTCTCGATATACCTGGCACAATCAGGAAGCAGACATTCATGG 735
 234 IaMetLeuLeuLysLeuArgGlnArgValLeuPheLeuLeuAspGlyTrp 250

|||||
 736 CCATGCTGCTGAAGCTGGCGCAGAGGCTTCTTCTTCTGATGCGCTAC 785
 251 AsnGluPheLysProGlnAsnCysProGluTrpIleGlnAlaLeuIleLysG 267
 |||||||
 786 AATGATTCAGAGCCCCAGAACTGCCAGAAATCGAAGCCCTGATTAAGGA 835
 267 uAsnHisArgPheLysAsnMetValIleValThrThrThrGluCysL 284
 |||||||
 836 AAACACCGGCTTCAAGAAACATGTCATGTCACACATCACCTAGATGCC 885
 284 euArgHisIleArgGlnPheGlyAlaLeuThrAlaGluValGlyAspMet 300
 |||||||
 886 TGAGGACATACGGCAGAGTTGTGTCCTGACTGCTGAGGTGGGATATG 935
 301 ThrGluAspSerAlaGlnAlaLeuIleArgGluValLeuIleLysGlu 317
 |||||||
 936 ACAGAACACAGCGCCCGAGCTCTCATCCGAGAACTGCTGATCAAGAGACT 985
 317 uAlaGluGlyLeuLeuLeuGlnIleGlnLysSerArgCysLeuArgAsn 334
 |||||||
 986 TGCTGAAGGCTGTGTGCTCCAAATTCAGAAATCCAGAGTGTGAAGATC 1035
 334 euMetLysThrProLeuPheValValIleThrCysAlaIleGlnMetGly 350
 |||||||
 1036 TCATGAAGACCCCTCTCTTGTGTCATCACTGTGCAATTCAGATGGCT 1085
 351 GluSerGluPheHisSerHisThrGlnThrThrLeuPheHisThrPheTy 367
 |||||||
 1086 GAAAGTAGTTCACACTCTCACACACAAACAAAGCTGTTCCTTACTTCTGA 1135
 367 rAspLeuLeuIleGlnLysAsnLysHisLysHisLysGlyValAlaAs 384
 |||||||
 1136 TGATCTGTGTATACAGAAACAAACAAACATTAAGGTGTGGCTGCAA 1185
 384 eArgPheIleArgSerLeuAsnHisCysGlyAspLeuAlaLeuGluGly 400
 |||||||
 1186 GTGACTTATTCGGAGCCTGGACCACTGTGGAGACTGACTGTGAGGGT 1235
 401 ValPheSerHisLysPheAspPheGluLeuGlnAspValSerSerValAs 417
 |||||||
 1236 GTGTTCTCCCAACAGTTGATTTCGAACCTGCAGATGTGTCACAGCTGAA 1285
 417 nGluAspValLeuLeuThrThrGlyLeuLeuCysLysTrpThrAlaGln 434
 |||||||
 1286 TGAGGATGTCTGCTGACAACTGGGCTCTCTGTAATATACACTGAAA 1335
 434 rGpPheLysProLysTrpTrpPhePheHisLysSerPheGlnGluTrpThr 450
 |||||||
 1336 GGTTCAAAGCCAAAGTATTAATTTCTTTCACAAAGTCAATTCAGAGACTACCA 1385
 451 AlaGlyArgArgLeuSerSerLeuLeuThrSerHisGluProGluGluVal 467
 |||||||
 1386 GCAGAGCAAGACATCAGACAGTTATTGACGTCTCATAGCCACAGAGAGGT 1435
 467 lThrLysGlyAsnGlyTrpLeuGlnLysMetValSerLysAspIleT 484
 |||||||
 1436 GACCAGAGGGAATGTACTTGCAGAAATGTGTTCCATTTCGAGACATTA 1485
 484 hrSerThrTrpSerSerLeuLeuArgTrpTrpCysGlySerSerValGlu 500
 |||||||
 1486 CATCCACTTATAGCAAGCTGCTCCGGTACACTGTGGGTCTCTGTGGAA 1535
 501 AlaThrArgAlaValMetLysHisLeuAlaAlaValTrpGlnHisGlyCys 517
 |||||||
 1536 GCCACACAGGCTGTATTGAAACACCTGCCACAGATGATTAACACGGCTG 1585
 517 sLeuLeuGlyLeuSerIleAlaLysArgProLeuTrpArgGlnGluSerL 534
 |||||||
 1586 CTTTCTCGAGATTTCATCCGCAAGAGGCTCTGTGGAGACAGGAATCTT 1635
 534 euGlnSerValLysAsnThrThrGluGlnGluIleLeuLysAlaIleAsn 550
 |||||||

1636 TCGAAGGTGTGAAAAACACCATGAGCAGAAATTCGAAAGCCATTAAC 1685
 551 ILeuSerPheValGluCysGlyIleHisLeuTyrGlnGluSerThrse 567
 1686 ATCAATTCCTTTGAGAGTGGCATCTTATATCAAGAGAGTACATC 1735
 567 rLySerAlaLeuSerGlnGluPheGlnAlaPhePheGlnGlyLysSerL 584
 1736 CAATATGAGCCCTGAGCCAGATTTGAAGCTTCTTCAGGTAAAGCT 1785
 584 eutyrlLeuSerGlyAsnIleProAspTyrLeuPheAspPhePheGlu 600
 1786 TATATATCAATCAGGAGACATCCCGATTACTATTTGACTTCTTGA 1835
 601 HisLeuProAsnCysAlaSerAlaLeuAspPheIleLysLeuAspPheTy 617
 1836 CATTGCGCCATTTGTGCAGAGTGGCTGAGACTGCTATTAAGTGGACTTTA 1885
 617 rGlyGlyAlaMetAlaSerTrpGluLysAlaAlaGluAspThrGlyLys 634
 1886 TGGGGGAGCTATGGCTTCATGGGAAAGGCTGCAGAGACAGAGGTGGA 1935
 634 LeuHisMetGluGlnAlaProGluThrTyrIleProSerArgAlaValSer 650
 1936 TCCACATGAGAAAGGCCCCAGAAACCTACATTCCTCCAGAGGCTGATCT 1985
 651 LeuPhePheAsnTrpLysGlnGluPheArgThrLeuGlnValThrLeuArg 667
 1986 TTGTTCTTCACTGGAAGAGAGATTCAGAGCTCTGGAGGTACACACTCCG 2035
 667 GAspPheSerLysLeuAsnLysGlnAspIleThrTyrLeuGlyLysIleP 684
 2036 GGAATTCAGCAAGTTGAAATGACAGATATCACATATCTGGGAAATAT 2085
 684 heserSerAlaThrSerLeuArgLeuGlnIleLysArgCysAlaGlyVal 700
 2086 TCAGCTCTGCACAAAGCCTCAGGCTGCATAATTAAGAGATGCTGGTGTG 2135
 701 AlAGlySerLeuSerLeuValLeuSerThrCysLysAsnIleTyrSerLe 717
 2136 GCTGGAAAGCTCAGTTGGTCTCAGCACCCTGTAAGACATTTATTTCTCT 2185
 717 uMetValGluAlaSerProLeuThrIleGluAspGluArgHisIleHis 734
 2186 CAGGTGGAAGCCAGTCCCTCAACATAGAAGATGAGAGCACATCATCAT 2235
 734 eTyValThrAsnLeuLysThrLeuSerIleHisAspLeuGlnAsnGlnArg 750
 2236 CTGTAAACAAACCTGAAAACTTGATATCATGATGACCTACAGAAATCAACG 2285
 751 LeuProGlyLysLeuThrAspSerLeuGlyAsnLeuLysAsnLeuThrLys 767
 2286 CTGCGCGGTGGTCTGACTGACAGCTTGGGTAACTGAAAGACCTTACAAA 2335
 767 sLeuIleMetAspAsnIleLysMetAsnGlnGluAspAlaIleLysLeu 784
 2336 GGCATTAATGATTAACATAAGATGAAGAAGATGCATATAAAGCTAG 2385
 784 laGluGlyLeuLysAsnLeuLysLysMetCysLeuPheHisLeuThrHis 800
 2386 CTGAAGGCCCTGAAAAACCTGAAGACATGTGTATTATTCTTTGACCAC 2435
 801 LeuSerAspIleGlyGlnGlyMetAspTyrIleValLysSerLeuSer 817
 2436 TTGTTCTGACATTTGAGAGGGAATGATTAATGATCAAGTCTCTGTCAAG 2485
 817 rGluProCysAspLeuGlnGluIleGlnLeuValSerCysCysLeuSer 834
 2486 TGAACCTGAGCCTGGAAGAATTAATTAATCTCCTGCTGTGCTG 2535
 834 laAsnAlaValLysIleLeuAlaGlnAsnLeuHisAsnLeuValLysLeu 850
 2536 CAAATGAGTGAATAATCTAGTTCAGAACTTCACAATTTGGTCAAACTG 2585

851 SerIleLeuAspLeuSerGluAsnTyrLeuGlnLysAspGlyAsnGlnAla 867
 2586 AGCATTCCTGATTTATCAGAAAAATTAACCTGGAAGAAAGATGAATAGAC 2635
 867 aLeuHisGluLeuIleAspArgMetAsnValLeuGlnGlnLeuThrAla 884
 2636 TCTTCAATGAACATGATCGACAGATGAACGTCGTAGAAGACGTACCGCAC 2685
 884 eukLeuProTrpGlyCysAspValGlnGlySerLeuSerSerLeuLeu 900
 2686 TGATGCTGCCCTGGGCTGTGAGCTCAAGGCGAGCTGAGACCCGTGTG 2735
 901 LysHisLeuGlnGluValProGlnLeuValLysLeuGlyLysLysAsnTr 917
 2736 AAACATTTGGAGGAGGTCCACAACTGTCACAGCTTGGGTGAAGAACTG 2785
 917 ParGLeuThrAspThrGluIleArgIleLeuGlnAlaPhePheGlyLys 934
 2786 GAGACTCAGAGATACAGAGATTAAGATTTAGTGCATTTTGGAAAGA 2835
 934 snProLeuLysAsnPheGlnGlnLeuAsnLeuAlaGlyAsnArgValSer 950
 2836 ACCCTCTGAAAAAATTCACAGAGTTGAATTTGGCGGAAATGCTGTGAGC 2885
 951 SerAspGlyTrpLeuAlaPheMetGlyValPheGluAsnLeuLysGlnLe 967
 2886 AGTGATGATGATGCTTCCCTTCATGGGTGTATTTGAAGAACTTAAAGCAAT 2935
 967 uValPhePheAspPheSerThrLysGlnPheLeuProAspProAlaLeuV 984
 2936 AGTGTCTTTTGAATTTAGTACTTAAGAATTTCTACTGATCCAGCATTTG 2985
 984 aArgLysLeuSerGlnValLeuSerLysLeuThrPheLeuGlnGlnAla 1000
 2986 TCAGAAACCTTACCAAGGTATTCACAGTTAACTTTCTGCAAGAGCT 3035
 1001 ArgLeuValGlyTrpGlnPheAspAspAspAspLeuSerValIleThrG 1017
 3036 AGGCTGTGGGTGGCAATTTGATGATGATGATCTCAGTGTATTATACAG 3085
 1017 yAlaPheLysLeuValThrAla 1024
 3086 TGCTTTAAACTAGTAAGTCT 3107
 seq_name: /SIDS2/ycgdata/geneseq/geneseqn/NA2001.DAT:AAH98254
 seq_documentation_block:
 ID AAH98254 standard; cDNA; 3545 BP.
 AC AAH98254;
 XX
 DT 12-OCT-2001 (first entry)
 XX
 DE Murine EST-derived coding sequence SEQ ID NO: 111.
 XX
 DE Human, sheep, pig, cow, fruit fly, yeast, hamster, macaque, horse;
 KW tomato, monkey, dog, sea urchin; expressed sequence tag; EST;
 KW diagnostics; forensic test; gene mapping; genetic disorder;
 KW biodiversity; gene therapy; nutrition; ss.
 OS Mus musculus.
 XX
 PN WO200154477-A2.
 XX
 PD 02-AUG-2001.
 XX
 PF 25-JAN-2001; 2001WO-US02687.
 XX
 PR 25-JAN-2000; 2000US-0491404.
 PR 17-JUL-2000; 2000US-0617746.
 PR 03-AUG-2000; 2000US-0631451.
 PR 15-SEP-2000; 2000US-0663870.

XX (HYSE-) HYSEQ INC.
PA Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI Cao Y, Drmanac RA, Zhang J, Werhman T;
XX WPI: 2001-476164/51.
DR P-PDB: AAM23595.
XX
XX Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use -
PS Claim 1: Page 250-251, 1275pp; English.

The present invention provides the protein and coding sequences of novel proteins from a variety of organisms, including human, dog, cat, horse, cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea urchin and tomato. These were derived from expressed sequence tags (ESTs) from the organism of interest. They can be used in diagnostics, CC forensics, gene mapping, identification of mutations, to assess CC biodiversity and for nutritional purposes. The present sequence is a CC of the invention.

XX
XX
SQ Sequence 3545 BP; 1038 A; 755 C; 816 G; 936 T; 0 other;

alignment_scores:
Quality: 5259.00 Length: 1024
Ratio: 5.151 Gaps: 0
Percent Similarity: 99.707 Percent Identity: 99.707

alignment_block:
US-09-697-089-2 x AAH98254 ..

Align seg 1/1 to: AAH98254 from: 1 to: 3545

1 MetLSPHeiLeuLYSASPASerArgAlaLeuIleGlnArgMetGlyLeu 17
232 ATGAATTTTCATTAAGACAAATAGCCGACCTTATTCAAAGAAATGGGAT 281
17 TThValIleuLYSGInIleThrAspAspLeuPheValIleThrAsnValLeu 34
282 GACTGTTTAAAGCAATCACAAGATGACCTTATTTGATGAAATGTTCTGCA 331
34 snArgGluGluValAsnIleIleCysGluLYSValGluGlnAspAla 50
332 ATCCGGAAGATGAACATCATTTGCTCGAAGAGTGAGACAGATGCT 381
51 AlaArgGlyIleIleHisMetIleLeuLYSGIleuLYSGIleuSerGlySerGly 67
382 GCTAGAGGATCATTCACATGATTTTGAAAAGGGTTCAGAGTCTCTGTA 431
67 nLeuPheLeuLYSSerLeuLYSGIleuThrAsnIleThrProLeuPheGlnAsp 84
432 CCTCTTTTAAATCCCTTAAGAGTGAAGTACTATCTTATTTGAGACT 481
84 euAsnGlyGlnSerLeuPheHisGlnThrSerGluGlyAspLeuAspAsp 100
482 TGAATGCAACAAGTCTTTTCATCAGACATCAGAAAGAGACTTGAGCAT 531
101 LeuAlaGlnAspLeuLYSAspLeuThyHisThrProSerPheLeuAsp 117
532 TTGGCTCGAGATTAAAGACTTGTACCATACCCCATCTTTTCTCAACT 581
117 eTyPProLeuGluGluAspIleAspIleIlePheAsnLeuLYSSerThr 134
582 TTATCCCTTGAGTGAATGATGACATTTATTTTAACTTGAAGAAGACCT 631
134 heThrGluProValLeuThrPArgLYSAspGlnHisHisArgValGlu 150
632 TCACAGAAACCTGCTCTGCGAAGAAAGCAACACCATCACCGCTGGAG 681
151 GlnLeuThrLeuAsnGlyLeuLeuGlnAlaLeuGlnSerProCysIleIle 167

682 CAGCTGACCTCGAATGGCTCTCGAGCTCTCAGAGCCCTCAGATCAT 731
167 eGlnGlyGluSerGlyLYSGIleuLYSSerThrLeuLeuGlnArgIleAla 184
732 TGAAGGGGAATCTGGCANAGCAAGTCCACTCTGCTGCGAGCTTGGCA 781
184 eLeuThrPglySerGlyLYSCysLYSAlaLeuThrLYSPheLYSPheVal 200
782 TGCTCTGGGGCTCCGAAAGTGAAGGCTGTGACCAAGTTCAATTCGTC 831
201 PhePheLeuArgLeuSerArgAlaGlnGlyGlyLeuPheGluThrLeu 217
832 TTCTCTCCCTCCAGCAGAGGCGCCAGGCTGACCTTTTGAACCCCTG 881
217 SASpGlnLeuLeuAspIleProGlyThrIleArgLYSGInThrPheMet 234
882 TGATCAACTCTCGATATACCTGGCACATAGAGACAGACATTCATG 931
234 IametLeuLeuLYSLeuArgGlnArgValLeuPheLeuAspGlyTyr 250
932 CCATGCTGCTGAAGCTGGCGCAGAGGGTTCTTCTCTGTGATGGCTAC 981
251 AsnGluPheLYSProGlnAsnCysProGluIleGlnAlaLeuIleLYSG 267
982 AATGAATTCAGCCCGACAGACTGCCAGAAATCGAAGCCGTGATPAAAGA 1031
267 uAsnHisArgPheLYSAsnMetValIleValIleThrThrGluCysL 284
1032 AAACCACCGCTTCAAGAAACATGTCATGTCACACATACCATGATGCC 1081
284 euArgHisIleArgGlnPheGlyAlaLeuThrAlaGluValGlyAspMet 300
1082 TGAGGACATATGCGCAGTGTGTCCTGACTGCTGAGTGCGGATATG 1131
301 ThrGluAspSerAlaGlnAlaLeuIleArgGluValLeuIleLYSGIle 317
1132 ACAGAAAGCAGCGCCAGGCTCTCATCGAAGATGCTGATCAAGAGACT 1181
317 uAlaGluGlyLeuLeuLeuGlnIleGlnLYSAspArgCysLeuArgAsn 334
1182 TGCtGAAGGCTGTGCTCCCAATTCAGAAATCCAGGTCCTTGAGGATC 1231
334 euMetLYSThrProLeuPheValIleThrCysAlaIleGlnMetGly 350
1232 TCATGAAGACCCCTCTGCTGTCATCACTGTGCAATCCAGATGGGT 1281
351 GluSerGluPheHisSerHisThrGlnThrThrLeuPheHisThrPhe 367
1282 GAAAGTGAAGTCCACTCTCACACACAAACAGCGCTGTTCCTTACTCT 1331
367 rAspLeuLeuIleGlnLYSAsnLYSHisLYSHisLYSGLYValAlaAs 384
1332 TGATCTGTTGATACAGAAAACAAACAAACATTAAGGTGGCTGCA 1381
384 eRAspPheIleArgSerLeuAspHisCysGlyAspLeuAlaLeuGlu 400
1382 GTGACTTCATTCGGAGCCTGGACCACTGTGATACCTGCTGAGAGGT 1431
401 ValPheSerHisLYSPheAspPheGluLeuGlnAspValSerSerVal 417
1432 GTGTTCCTCCCAAGATTGATTTGAACTGCAAGATGTGTCCACAGGAA 1481
417 nGluAspValLeuLeuThrThrGlyLeuLeuCysLYSThrAlaGln 434
1482 TGAGGATGCTCTGCTGACAACTGGGCTCTGTAAATATACACTGAAA 1531
434 rPheLYSProLYSTyLYSPhePheHisLYSSerPheGlnGluTyrThr 450
1532 GCTTCAAGCCCAAGATTAATTTTTCACAAAGTCATTCAGAGTAAACA 1581
451 AlaGlyArgGluSerSerLeuLeuThrSerHisGluProGluGluVal 467

1582 GCAGGACGACACTCAGAGTTTATGACGTCTCATGACGACGAGAGCT 1631
 467 LThrLysGlyAsnGlyTyrLeuGlnLysMetValSerIleSerAspIleT 484
 1632 GACCAAGGGAGATGTTACTTGCAAAAATGTTCCATTTCGAGCATTA 1661
 484 hrSerThrTyrSerSerLeuLeuArgTyrThrCysGlySerSerValGlu 500
 1682 CATCCACTTATAGCAGCCTGCTCGGTACACCTCGGTGTCATCTGTGGAA 1731
 501 AlaThrArgAlaValMetLysHisLeuAlaAlaValTyrGlnHisGlyC 517
 1732 GCCACCGAGGCTGTTATGAAGCACCTCGCAGCAGTGTATCAACACGCGTG 1781
 517 sleuLeuGlyLeuSerIleAlaLysArgProLeuTyrArgGlnLysSer 534
 1782 CTTCTCGAGCTTTCATCGCCAAAGAGCCCTCTCTGAGACAGGAAACCTT 1831
 534 euGlnSerValLysAsnThrThrGlnGlnGlnLileuLysAlaIleAsn 550
 1832 TGCAAAGTGTGAAAAACACCACTGACAGCAAGAAATTCGAAAGCCATTAAC 1881
 551 IleAsnSerPheValGluCysGlyIleHisLeuTyrGlnLysSerThrSe 567
 1882 ATCAATTCTTTGTAGAGTGTGCATCTTATATCAAGAGACTACATC 1931
 567 rLysSerAlaLeuSerGlnGlnPheGlnAlaPhePheGlnGlyLysSer 584
 1932 CAAATWCAGCCCTGAGCCAAAGAAATTTGAAGCTTCTTCAAGGTAAAGCT 1981
 584 euTyrIleAsnSerGlyAsnIleProAspTyrLeuPheAspPheGlu 600
 1982 TATATATCAACTCGAGGACATCCCGATTACTTATTTGACTTCTTGAA 2031
 601 HisLeuProAsnCysAlaSerAlaLeuAspPheIleLysLeuAspPheTyr 617
 2032 CATTTGCCAATTGTGCCAGTGTCTGACTCATTTAACTGGGCTTTTAA 2081
 617 rGlyGlyAlaMetAlaSerTyrGlyLysAlaAlaGlnAspThrGlyGlyT 634
 2082 TGGGGAGCTATGGCTTCATGAGGAAAGGCTGCAGAGACACAGGTGAA 2131
 634 leHisMetGlnGluAlaProGlnuThrTyrIleProSerArgAlaValSer 650
 2132 TCCACATGAGAAAGGCCCGCAAAACCTACATTCACAGAGGCGCTGTCT 2181
 651 LeuPhePheAsnTyrPylsGlnLupheArgThrLeuGlnuValThrLeuAr 667
 2182 TTGTTCTTCAACTGCAAGCAGGAATTCAGACTCTGAGGCTCACACTCCG 2231
 667 gAspPheSerLysLeuAsnLysGlnAspIleThrTyrLeuGlyLysIleP 684
 2232 GGATTTTACGAAAGTTGAATAGCAAGATATCAGATATCTGGGGAAATAT 2281
 684 heSerSerAlaThrSerLeuArgLeuGlnIleLysArgCysAlaGlyVal 700
 2282 TCAGCTCTGCCACAAAGCTCAGGCTGCAAAATAAAGATATGCTGCTGCTG 2331
 701 AlaGlySerLeuSerLeuValLeuSerThrCysLysAsnIleTyrSerLe 717
 2332 GCTGGAAGCCCTCAGTTTGTCTCTCAGACCTGTAAAGAAATTTTCTCT 2381
 717 uMetValGluAlaSerProLeuThrIleGlnuAspGluArgHisIleThrS 734
 2382 CATGTGTGAAGCCAGTCCCTCACCATTAGAAAGAGAGGACACATCATCAT 2431
 734 erValThrAsnLeuLysThrLeuSerIleHisAspLeuGlnAsnGlnArg 750
 2432 CTGTAAACAAACCTGAAACCTTGAGTATTTCATGCTACGAAATCAACG 2481
 751 LeuProGlyLysLeuThrAspSerLeuGlyAsnLeuLysAsnLeuThrL 767
 2482 CTGCGGGGTGTGTGACTGACAGCTTGGGTAACTTGAAGAACTTACAAA 2531

767 sleuIleMetAspAsnIleLysMetAsnGlnGluuAspAlaIleLysLeuA 784
 2532 GCTCATATATGATTAACATAAAGATGAATGAAGATGCTATTAACCTAG 2581
 784 lacGluGlyLeuLysAsnLeuLysLysMetCysLeuPheHisIleuThrHis 800
 2582 CTGAAGGCCCTGAAAAACCTGACAGACATGCTTATTTCATTGACCCAC 2631
 801 LeuSerAspIleGlyGlyMetAspTyrIleValLysSerLeuSerSe 817
 2632 TTGTCTGACATTGTGAGAGGGGAATGATTAATATGATCAAGCTCTGCAAG 2681
 817 rGluProCysAspLeuGlnGlnIleGlnLeuValSerCysCysLeuSerA 834
 2682 TGAACCTGTGACCTTGAAGAAATTCATTAATTAATGCTCTGCTGTCTTG 2731
 834 laAsnAlaValLysIleLeuAlaGlnAsnLeuHisAsnLeuValLysLeu 850
 2732 CAATGCGAGTGAATAATCCTAGCTCAGAAATCTTCAATTTGCTCAAACTG 2781
 851 SerIleLeuAspLeuSerGluAsnTyrLeuGlnLysAspGlyAsnGlnuAl 867
 2782 AGCATCTTGTATTATCAGAAATTAATCTGCAAAAGATGGAATGAAGAC 2831
 867 alaHisGlnLeuIleAspArgMetAsnValLeuGlnGlnLeuThrAla 884
 2832 TCTTCATGAACCTGATGACAGATGAACGCTGTAGAACACCTCACCGCAC 2881
 884 euMetLeuProTyrPylsCysAspValGlnGlySerLeuSerSerLeuLeu 900
 2882 TGATGCTGCCCTGGGCTGTGAGCTGCAAGGACCTGACACGCTGTG 2931
 901 LysHisLeuGlnGluValProGlnLeuValLysLeuGlyLeuLysAsnTr 917
 2932 AAACATTTGGAGAGGCTCCCAACTGCTCAACGCTTGGCTGAAAACTG 2981
 917 PArgLeuThrAspThrGlnuIleArgIleLeuGlyAlaPhePheGlyLysA 934
 2982 GAGACTCACGATACAGATGATTAATTTAGGTGCAATTTTGGAAAAA 3031
 934 snProLeuLysAsnPheGlnGlnLeuAsnLeuAlaGlyAsnArgValSer 950
 3032 ACCCTCGAAAAAATCTCCAGCACTGATTAATTTGGCGGGAATGCTGTGAC 3081
 951 SerAspGlyTyrLeuAlaPheMetGlyValPheGlnAsnLeuLysGlnLe 967
 3082 AGTGATGAGATGCGCTTCCCTCATGGGTGTATTTGAAGAACTTTAAAGCAATT 3131
 967 uValPhePheAspPheSerThrLysGlnLupheLeuProAspProAlaLeu 984
 3132 AGTGTTTTTGGACTTTATTAAGAAATTTTACCTGATCCGCAATTTAG 3181
 984 alArgLysLeuSerGlnValLeuSerLysLeuThrPheLeuGlnGlnuAla 1000
 3182 TCAGAAAACTTACCAAGTGTATCCAAAGTTACTTTCTGCAAGAACT 3231
 1001 ArgLeuValGlyTyrPylsPheAspAspAspAspLeuSerValIleThrG 1017
 3232 AGGCTTGTGCTGGCAATTTGATGATGATGATGATCAGTGTATTATACAG 3281
 1017 yAlaPheLysLeuValThrAla 1024
 3282 TGCTTTTAAACTAGTAACCTGCT 3303

seq_name: /SIDS2/gcdata/geneseq/geneseqn/NA2001.DAT: AAS03946

seq_documentation_block:

ID AAS03946 standard: DNA: 3615 BP.

XX AAS03946;

AC AAS03946;

XX AAS03946;

DT 12-SEP-2001 (first entry)

XX	Human caspase recruitment domain 12 (CARD-12) genomic DNA.
DE	
XX	
KW	Caspase recruitment domain; CARD-12; apoptosis; stress-related pathway;
KW	cancer; viral infection; poxvirus; adenovirus; autoimmune disorder;
KW	systemic lupus erythematosus; arthritis; neurological disorder; stroke;
KW	Alzheimer's disease; amyotrophic lateral sclerosis; hematologic disease
KW	aplastic anemia; myocardial infarction; inflammatory disorder;
KW	Crohn's disease; insulin-dependent diabetes; contact dermatitis;
KW	psoriasis; graft rejection; bacterial infection; lepromatous leprosy;
KW	tuberculosis; ischaemic brain injury; hypoxic brain injury; ds;
KW	kidney ischaemia; reperfusion injury; acute bacterial meningitis;
KW	excitotoxic brain damage; liver disease.

FF	Key	Location/Qualifiers
FT	CDS	1..3615
FT		/*tag= a
FT		/product= "Human CARD-12"
PN		WO200130971-A2.
PN		
PD		03-MAY-2001.
XX		
PF		26-OCT-2000; 2000WO-US29643.
XX		
PR		27-OCT-1999; 99US-0161822.
XX		
PA		(MILL-) MILLENNIUM PHARM INC.
XX		
PI		Bertlin J, Robison KE;
XX		
DR		WPI, 2001-308628/32.
DR		P-PSDB; AA002881.

PT Isolated caspase recruitment domain-12 polypeptide and nucleic acids
PT encoding them, useful for treating and diagnosing disorders associated
PT with abnormal apoptosis such as cancer, arthritis and Alzheimer's
PT disease -
XX
PS Disclosure, Fig 2, 93pp, English.

The sequence represents a genomic DNA which encodes the human caspase recruitment domain 12 (CARD-12) polypeptide. CARD domains are found in a number of proteins that transmit signals that activate apoptosis and inflammatory pathways in response to stress and other stimuli. Therefore, CARD-12 and its corresponding nucleic acid may be used in treatment and diagnosis of patients suffering from disorders associated with an abnormal level (an increase or a decrease) of apoptotic cell death or abnormal activity of stress-related pathways. The disorders include cancer, viral infections (e.g. caused by poxviruses, adenoviruses), autoimmune disorders (e.g. systemic lupus erythematosus, arthritis), neurological disorders (e.g. Alzheimer's disease, amyotrophic lateral sclerosis), haematological diseases (e.g. aplastic anaemia, myocardial infarction, stroke), inflammatory and immune system disorders (e.g. Crohn's disease, insulin-dependent diabetes, contact dermatitis, psoriasis, graft rejection), bacterial infections (e.g. tuberculosis, leprosy), ischaemic and hypoxic brain injury, kidney ischaemia/reperfusion injury, excitotoxic brain damage, acute bacterial meningitis and liver disease.

SQ Sequence 3615 BP; 1041 A; 811 C; 845 G; 918 T; 0 other;

```
alignment_scores:
  Quality: 5214.00
  Ratio: 5.137
  Gaps: 1
Percent Similarity: 98.162
Percent Identity: 98.066
```

alignment_block:
US-09-697-089-2 x AAS03946

Align seg 1/1 to: AAS03946 from: 1 to: 3615

1	MeAspSphelIeLyspSpsnSerArgAlaLeuIleGlnArgMeGI	17
418	GTAAATTTATTAAGACACATATAGCCAGCCCTTATTACAAAGATGGAA	467
17	lThrValIleLysGlnIleThrAspSLeuPheValTrpAsnValLeu	34
468	GACTGTATAAAGCAATATACAGATGACCTATTGTATGCAATGTTTCGA	517
34	snArgIuGIuValAsnIleIleCysCysGluValGluGlnAspIa	50
518	ATGCGCAAGAAAGTAAACATCATTTGCTGGCAAGAGTGGAGGAGATGCT	567
51	AlaArgGlyIleIleHisMetIleLeuLysGlySerGluSerCysAs	67
568	GCTAGAGGGATCATTTACATGATTTTGAAGAAAGGGTTGAGATCTCTGTA	617
67	nLeuPheLeuLysSerLeuLysGluTrpAsnTyrProLeuPheGlnAsp	84
618	CCCTCTTCTTAAATCCCTTAAGAGTGAACATCCTCTATTTCAGAGCT	667
84	euaSngIysInser	88
668	TGATGACAAAGTTTGGAGACACAGAAATTTGGGTCTTCTTAACATC	717
89LeuPheHisGlnThrSerGluGlyAspLeuAs	99
718	ACCTCTTCTCTAATAGCTCTTTTTCATCAGACATCAGAAAGAGACTTGG	767
99	AspLeuAlaGlnAspLeuLysAspLeuTyrHisThrProSerPheLeu	116
768	CGATTTGGCTCAGGATTTAAAGACTTGACCATACCCCATCTTTTGCA	817
116	snPheTyrProLeuGlnGlyAspIleAspIleIlePheAsnLeuLysSer	132
818	ACTTTTATCCCTTGGTGTAAGATATTGACATTATTTTAACCTGAAAGC	867
133	ThrPheThrGluProValLeuTrpArgLysAspGlnHisHisIAspVal	149
868	ACCTTCACGAACCTGTCTCTGGGAGGAAGGACCAACACATCACCCGGT	917
149	IleGluInLeuThrLeuAsnGlyLeuLeuGlnAlaLeuGlnInserProCysI	166
918	GAGACAGCTGACCTGAAAGGGCTCTCGAGAGGCTTTCAGAACCCCTCA	967
166	IleIleGluGlnLysSerGlyLysGlyLysSerThrLeuLeuGlnArgIle	182
968	TCATTTCMAAGGGAACTGTGGCAAAAGGCAAGTCCATCTCTCGACGCAAT	1017
183	AlaMetLeuTrpGlySerGlyLysCysLysAlaLeuThrLysPheLysPh	199
1018	GCCATCTCTGGGGCTCCGGAAAGTGCAGAGGCTCTGACCAAGTTCAAT	1067
199	eValPhePheLeuArgLysSerArgAlaGlnGlyLysLeuPheGluTrpL	216
1068	CGCTCTTCTCCGCCGCTCAGACAGAGGCCAGAGGTGAGACTTTTGAACCC	1117
216	eucYaspGlnLeuLeuAspIleProGlyThrIleArgLysGlnThrPhe	232
1118	TCGTGTATCACTCCGTGATTAACCTGGCACATAGGAACAGACATTC	1167
233	MetAlaMetLeuLeuLysLeuArgGlnArgValLeuPheLeuLeuAspG	249
1168	ATGCGCATCTGCTGAGCTGGCGCAGAGGGTCTTTTCTTCTTGATGG	1217
249	yTrpAsnGluPheLysProGlnAsnCysProGluIleGluAlaLeuIle	266
1218	CTACATGTAATTCAGGCCCAAGACTGCCAGAAATTCAGAGCCGTGATA	1267
266	ysGluAsnHisArgPheLysAsnMetValIleValThrThrThrThrGlu	282
1268	AGGAAACACCGCTTCAAGAACATGATCTGTCATCCACATCAACATGAG	1317

283 CysLeuArgHisIleArgInpHeGlyAlaLeuThrAlaGluValGlyAs 299
|||||
1318 TGCCTGAGGACATACGGCAGTTGGCCCTTACCTGCTAGGTGGGGCA 1367
299 pMetThrGluAspSerAlaGlnAlaLeuIleArgGluValIleLysG 316
|||||
1368 TATGACAGAGACAGCGCCAGGCTCTCATCCGAGAGTGTGATCAAG 1417
316 IuLeuAlaGluGlyLeuLeuLeuGlnIleGlnLysSerArgCysLeuArg 332
|||||
1418 AGCTTGCTGAAGCTTGTCTCCAAATTCAGAAATCCAGGTCTTGAGC 1467
333 AsnLeuMetLysThrProLeuPheValIleThrCysAlaIleGlnMe 349
|||||
1468 AATCTCATGAGACCCCTCTTGTGGTGTATCATCTTGTCATTCAGAT 1517
349 tGlyGluSerGluPheHisSerHisThrGlnThrThrLeuPheHisThrP 366
|||||
1518 GGGTGAAGTGAAGTCCACTCTCACACACAAACAGCTTCCATACCT 1567
366 heTyrrAspLeuLeuIleGlnLysAsnLysHisLysHisLysGlyValAla 382
|||||
1568 TCTATGATCTGTGATACAGAAAAACAAACAAACATAAAGCTGTGGCT 1617
383 AlaSerAspPheIleArgSerIleuAspHisCysGlyAspLeuAlaLeuG 399
|||||
1618 GCATGTACTTCACTTGGAGCTGGACCTGGACCTGGAGACCTAGCTCGGA 1667
399 uGlyValPheSerHisLysPheAspPheGluLeuGlnAspValSerSerV 416
|||||
1668 GGGTGTGTCTCCACAAAGTTGATTTGCAACTGACGAGATGTGCCAGC 1717
416 alaGlnGluAspValLeuLeuThrThrGlyLeuLeuCysLysTyrrThrAla 432
|||||
1718 TGAATGAGATGTCCTGCTGACACAGTGGCTCTCTGTAATATACAGCT 1767
433 GlnArgPheLysProLysTyrrLysPhePheHisLysSerPheGlnLys 449
|||||
1768 CAAGGTTCAGCCCAAGTAAATCTTTACACAGTCAATCCAGGAGTA 1817
449 rThrAlaGlyArgArgLeuSerSerLeuLeuThrSerHisGluProGluG 466
|||||
1818 CACAGCAGGACGAAGACTCAGAGTTATGAGCTCTCATGAGCCAGAG 1867
466 IuValIleThrLysGlyAsnGlyTyrrLeuGlnLysMetValSerIleSerAsp 482
|||||
1868 AGGTGACCAAGGGGAATGTTACTTGCAGAAAAATGGTTCCATTTGGGAC 1917
483 IleThrSerThrTyrrSerSerLeuLeuArgTyrrThrCysGlySerSerVa 499
|||||
1918 AATACATCCCTTATACAGACCTGCTCCGGTACACCTGTGGGTCACTGT 1967
499 IguAlaThrArgAlaValMetLysHisLeuAlaAlaValTyrrGlnHisG 516
|||||
1968 GGAAGCCACACAGGGCTGTTATGAAACACCTCCGACAGTGTATCAACAG 2017
516 IyGysLeuLeuGlyLeuSerIleAlaLysArgProLeuThrPargGlnGlu 532
|||||
2018 GGTGCTTCTCGGACTTTCATCGCCACAGAGGCTCTCGAGACAGGAA 2067
533 SerLeuGlnSerValLysAsnThrThrGluGlnGluIleLeuLysAlaI 549
|||||
2068 TCTTTCGCAAAAGTGA AAAACACACTGAGCAAGAAATTTCTGAACCAT 2117
549 eAsnIleAsnSerPheValGluCysGlyIleHisLeuTyrrGlnGluSerT 566
|||||
2118 AAACATCAATTTCTTTGTAGAGTGTGCAATTCATTTATATCAAGAGAGTA 2167
566 hTrSerLysSerAlaLeuSerGlnGluPheGlnAlaPhePheGlnGlyAs 582
|||||
2168 CATCAAAATACGCCCTGAGCCAGAAATTTGAAGCTTTCTTCAAGGTAA 2217

583 SerLeuTyrrIleAsnSerGlyAsnIleProAspTyrrLeuPheAspPheP 599
|||||
2218 AGCTTATATATCACTACAGGAACATCCCCGATTTACTTATTTGACTTCTT 2267
599 eGluHisLeuProAsnCysAlaSerAlaLeuAspPheIleLysLeuAsp 616
|||||
2268 TGAACATTTGGCCCAATTTGTCAAGTGCCTTGAGCTCATTTAACTGGACT 2317
616 heTyrrGlyGlyAlaMetAlaSerTPGluLysAlaAlaGluAspThrGly 632
|||||
2318 TTTATGGGGAGCTATGCTTCATGGGAAAGCTGCACAGAACACAGGT 2367
633 GlyIleHisMetGluGluAlaProGluThrTyrrIleProSerArgAlaVa 649
|||||
2368 GGAATCCACATGAGAGAGGCCCCAGAAACCTTACATTTCCAGCAGGGCTGT 2417
649 IserLeuPhePheAsnTrpLysGlnGluPheArgThrLeuGluValThrL 666
|||||
2418 ATCTTGTTCTTCAACTGAGAGAGGAATTCAGAGCTGTGAGGTCCACAC 2467
666 euArgAspPheSerLysLeuAsnLysGlnAspIleThrTyrrLeuGlyLys 682
|||||
2468 TCCGGGATTTCAAGCAAGTTGAATTAAGCAAGATATCAGATATCTGGGAAA 2517
683 IlePheSerSerAlaThrSerLeuArgLeuGlnIleLysArgCysAlaG 699
|||||
2518 AATTTACGCTTGCCACAGCTTCAGGCTGCAGAAATTAAGATGTGCTGG 2567
699 yValAlaGlySerLeuSerLeuValLeuSerThrCysLysAsnIleTyrrS 716
|||||
2568 TGTGGCTGAGAGCTCAGTTGTGCTCCTCAGCACCTGATGAAGACATTTAT 2617
716 erLeuMetValGluAlaSerProLeuThrIleGluAspGluArgHisIle 732
|||||
2618 CTCTCATGTGTGAAGCCATCCCTCACCATGAAAGATGAGGACATAC 2667
733 ThrSerValThrAsnLeuLysThrLeuSerIleHisAspLeuGlnAsnG 749
|||||
2668 ACATCTGTAAACAACCTGAAAACCTTGAGTATTCATGACTTACAGATCA 2717
749 nArgLeuProGlyGlyLeuThrAspSerLeuGlyAsnLeuLysAsnLeuT 766
|||||
2718 ACGGCTCGCGGGTGTCTGACTGACAGCTTGGTGAATCTGGAAGAACCTTA 2767
766 hTrLysLeuIleMetAspAsnIleLysMetAsnGluGlnAspAlaIleLys 782
|||||
2768 CAAAGCTCATTAATGATGAATAAGATGAATGAAGAAATGCTATAAAA 2817
783 LeuAlaGluGlyLeuLysAsnLeuLysLysMetCysLeuPheHisLeuTh 799
|||||
2818 CTAGCTGAAGGCTGAAAAACCTGAAAGAAATGTGTTTATTTTCATTTGAC 2867
799 rHisLeuSerAspIleGlyGluGlyMetAspTyrrIleValLysSerLeuS 816
|||||
2868 CCACCTTGCTGACATTTGAGAGGAAATGATGATTAATGACATGCTCTGT 2917
816 erSerGluProCysAspLeuGluGluIleGlnLeuValSerCysLeu 832
|||||
2918 CAAGTGAACCTGTGACCTTGAAGAAATTCATTAATGTCCTCGCTGTG 2967
833 SerAlaAsnAlaValLysIleLeuAlaGlnAsnLeuHisAsnLeuValLy 849
|||||
2968 TCTGCAAAATGCAAGTGAATTTCTAGCTGCAATTTCTTCAAAATTTGCTCA 3017
849 sLeuSerIleLeuAspLeuSerGluAsnTyrrLeuGluLysAspGlyAsnG 866
|||||
3018 ACTGAGCATTTCTTGAATTTATCAGAAAAATTAACCTGGA AAAAGTGAATG 3067
866 IuAlaLeuHisGluLeuIleAspArgMetAsnValLeuGlnGlnLeuThr 882
|||||
3068 AAGCTCTTATGAACTGATGACAGAGATGAACGTCTGAAACAGCTCAC 3117
883 AlaLeuMetLeuProTrpGlyCysAspValGlnGlySerLeuSerSerIle 899

3118	GCACGTGATGCTGGCCCTGGGGCTGTGAAGTGCAGGACGCGTGGACAGCCT	3167	
899	uLeuLysHISLeuGluGluValProGlnLeuValLysLeuGlyLeuLysA	916	
3168	GTTGAACAATTGGAGAGAGGTCCACAACTGTCACAGCTGGGTGAAAA	3217	
916	snTPRArgLeuThrAspThrGluIleArgGlyLeuGlyAlaPhePheGly	932	
3218	ACTGGAGACTCACAGATTACAGAAATTAGATTATTTAGTGCACTTTTGGGA	3267	
933	LysAsnProLeuLysAsnPheGlnGlnLeuAsnLeuAlaGlyAsnArgVa	949	
3268	AAGAACCCCTCTGAAAAAATCTCCAGCAGTTGAATTTGGCGGAAATCGNGT	3317	
949	LSerSerAspGlyTrpLeuAlaPheMetGlyValPheGluAsnLeuLysG	966	
3318	GAGCAGTGTATGATGGCTTCCTTCATGGGGTGATTTGGAGAAATTTTAAGC	3367	
966	IuLeuValAlaPhePheAspPheSerThrLysGluPheLeuProAspProla	982	
3368	AATTAGTGTATTTTGACTTTAGACTTAGACACTAAAGAAATTTTACCTGCATCCACA	3417	
983	IeuValArgLysLeuSerGlnValLeuSerLysLeuThrPheLeuGlnG	999	
3418	TTAGTCAGAAAACTTAGCCAAAGGTATATCCACAGTTACTTTTCGCAGA	3467	
999	uAlaArgLeuValGlyTrpGlnPheAspAspAspAspLeuSerValIleT	1016	
3468	ACCTAGACCTTTGGGTGGCCAAATTTGATGATGATGATCTCAGTGTATTA	3517	
1016	hr 1016		
3518	CA 3519		
seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT.AAH99561			

seq. documentation block:	seq. documentation block:
ID	ID
AAH99581 standard; cDNA; 2950 BP.	AAH99581 standard; cDNA; 2950 BP.
AAH99581:	AAH99581:
16-OCT-2001 (first entry)	16-OCT-2001 (first entry)
Human protein encoding cDNA sequence SEQ ID NO:416.	Human protein encoding cDNA sequence SEQ ID NO:416.
Human; cancer; ulcer; HIV infection; human immunodeficiency virus; antiinflammatory; antirheumatic; antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac; central nervous system; virucide; anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia; antipregnant; haemostatic; vulnery; antileucor; osteopathic; eczema; dermatologic; antiallergic; antisthmatic; antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic; antiparkinsonian; infection immunostimulant; gene therapy; antisense therapy; vaccine; inflammation antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis; cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity; genetic disease; haematopoietic disorder; platelet disorder; asthma; thrombocytopaenia; osteoporosis; severe combined immunodeficiency; allergic rhinitis; diabetes; multiple sclerosis; depression; Alzheimer's disease; Parkinson's disease; neurodegenerative disorder; neurological disorder; ss.	Human; cancer; ulcer; HIV infection; human immunodeficiency virus; antiinflammatory; antirheumatic; antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac; central nervous system; virucide; anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia; antipregnant; haemostatic; vulnery; antileucor; osteopathic; eczema; dermatologic; antiallergic; antisthmatic; antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic; antiparkinsonian; infection immunostimulant; gene therapy; antisense therapy; vaccine; inflammation antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis; cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity; genetic disease; haematopoietic disorder; platelet disorder; asthma; thrombocytopaenia; osteoporosis; severe combined immunodeficiency; allergic rhinitis; diabetes; multiple sclerosis; depression; Alzheimer's disease; Parkinson's disease; neurodegenerative disorder; neurological disorder; ss.
Homo sapiens.	Homo sapiens.
MO200153455-A2.	MO200153455-A2.
26-JUL-2001.	26-JUL-2001.
22-DEC-2000; 2000MO-US35017.	22-DEC-2000; 2000MO-US35017.
23-DEC-1999; 99US-0471275.	23-DEC-1999; 99US-0471275.
21-JAN-2000; 2000US-0488725.	21-JAN-2000; 2000US-0488725.
25-APR-2000; 2000US-0552317.	25-APR-2000; 2000US-0552317.

PA	(HYSE-) HYSEQ INC.
XX	
XX	Tang YT, Liu C, Dzmanac RT;
XX	WPI; 2001-457603/49.
DR	P-PSDB; AAM25640.
XX	
PT	Isolated human polynucleotides encoding polypeptides, useful for the
PT	treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
XX	
PS	Claim 1; Page 511-512; 1217bp; English.
XX	
XX	AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
CC	AAH25963. The proteins can have activities based on the tissues and
CC	cells they are expressed in, such as: antinflammatory; antirheumatic;
CC	antiallergic; immunosuppressive; antibacterial; endocrine; cardiant;
CC	central nervous system; viral; anti-HIV; fungicide; antitumagen;
CC	cardiovascular; antianaemic; antiagregant; haemostatic; vulnerary;
CC	antitumor; osteoprotic; dermatological; antiallergic; antiasthmatic;
CC	antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
CC	antiparkinsonian; and immunostimulant. The proteins and polynucleotides
CC	encoding them can be used in gene therapy, antisense therapy and vaccine
CC	production. The proteins and polynucleotides are useful for screening for
CC	agonists or antagonists of a protein and for the treatment and diagnosis
CC	of disorders associated with the activity of a protein e.g. inflammation,
CC	rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC	neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC	infections, autoimmunity, genetic diseases, haematopoietic disorders,
CC	anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
CC	osteoporosis, severe combined immunodeficiency, eczema, allergic
CC	rinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC	Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC	neurological disorders.
XX	
XX	Sequence 2950 BP; 915 A; 592 C; 628 G; 815 T; 0 other;

```

Alignment_scores:      Length: 642
Quality: 3266.00
Ratio: 5.111
Gaps: 0
Percent Similarity: 99.553    Percent Identity: 99.377

Alignment_block:
US-09-697-089-2 x AAH9581 ..

Align seg 1/1 to: AAH9581 from: 1 to: 2950

363 AAsersApheIIeArgSerLeuAspIscysGlyAspLeuAlaLeuG1 399
|||||
2 GCAAGTACTTCATTCGGAGGCTGGACCACTGGTAATCATTCGTGGA 51
|||||

399 uGlyAlpheSerHisLysPheaspPheGluLeuG1AspValSerSerV 410
|||||
52 GGGTGTGTTCCCCACAGTTTGATTGCACTGAGAGTGTGCCAGC 101
|||||

416 aAsnGluAspValLeuLeuThrThrGlyLeuLeuCysLysTyrThAla 433
|||||
102 TGAATGAGAGATGCTCTGTGCAACTGGGCTCTCTGTAATAATACAGCT 151
|||||

433 GluArgPheLysProLysTyrLysPhePheHisLysSerPheGlnGluTy 449
|||||
152 CAAAGGTCMAAGCCAAAGTATTAATCTTTTACACAGTCATCCAGGATA 201
|||||

449 rThraIaGlyArgArgLeuSerSerLeuLeuThrSerIscIaProGluG 466
|||||
202 CACACGACGAGCAAGAACTCACACAGCTTATTCAGCTCATGACCAAGC 251
|||||

466 luValThrLysGlyAsnGlyTyrLeuGlnLysMetValSerIleSerAsp 483
|||||
251 AGTGTACCAAGGGGAAATGTTACTTGCAGAAATGGTTTCCATTTGGAC 301
|||||

483 IleThrSerThrTyrSerSerLeuLeuArgTyrThrCysGlySerSerVa 499

```

```

|||||
302 ATTACATCCACTTATAGACAGCCTGCTCCGGTACACCTGTGGTCACTCTGT 351
499 IGLUALIATHARGALAYALMELTYSHSLEUALAALAYALTYYRGLNHISC 516
352 GGAAAGCCACAGGGCTGTTATGAAGACCTCCGACAGATGTATCAACACG 401
516 IYCSYLEUENUGLYLEUSERILEALAYSARPROLEUTPRARGINGLU 532
402 GCGGCTTCTCGGACTTTCATCGCCAGAGGCTCTCTGGAGACAGGA 451
533 SERLEUGINSEVALLYSANTHRHGLINGLUILEULYSALAI 549
452 TCTTGCAAAAGTGTGAAAAACACCTGAGCAAGAAATTCGAAACCAT 501
549 EASNIIEANSERPHEVALGLUCYSELYILEHISLEUTYRGLINLUSERT 566
502 AAACATCAATTCCTTGTAGAGTGTGCATCCATTTATATCAAGACAGTA 551
566 HRSERYSERALALEUSERGLINGLUPhGLUALAPHEPHEGLINGLYLS 582
552 CATCCAAATCAGCCCTGAGCCAAAGAAATTTGAGCTTCTTTCAAGGTAA 601
583 SERLEUTYRIEASNSERGLYASNIIEPROASPTYRLEUPHEASPHEPH 599
602 AGCTTATATATCAACTCAGGAGAACATCCCGATTACTTATTGACTTCTT 651
599 EGLUHSLEUPROASNCYSALASERALEUASPHEIIELYSLEUASPP 616
652 TGAACATTTTGCCCAATTTGTCAAGTCTGTGGACTTCATTAAATCGGGCT 701
616 HETRYGLYGLYALAMETALASERTIPGLULYSALALAGIUSAPTRHGLY 632
702 TTTATGGGGAGCTATGGCTTCTATGGGAAAAGGCTCGAGAGACAGGT 751
633 GLYILEHISMETGLULUALPROGLUHTYRIIEPROSEARGLAYA 649
752 GGAATCCACATGAGAGAGGCCCAAGAACCTTACATTCGACGAGGGCTGT 801
649 ISERLUPHEPHEASNTYRGLINGLUPHEATGTHLEUGLUALIATHL 666
802 ATCTTTGTCTTCACTGAGAGCAGAAATTCAGGACTCTGGAGGTCAAC 851
666 EULARGASPHESELYSLEUASNLYSGLINASPILETHRTRYLEUGLYLS 682
852 TCCGGGATTTACAGCAAGTTGAATTAACAGCATATTCAGATATCTGGGAAA 901
683 IIEPHESESERALATHRSERLEUARGLEUGNIIELYSARGYSALAI 699
902 ATATTCAGCTCTGCCACAAGCCTCAGGCTGCAGCAATTAAGAGATGCTGG 951
699 YVALIALAGLYSERLEUSERLEUVALLEUSERTHRCYLSYASNIIEYTS 716
952 TGTGGCTGGAGGCTCAGTTGTCTCTCAGCACCTGTAAAGAACATTATTT 1001
716 ERLEUMETVALIGUALASERPROLEUTHRIEGLUASPGUARGHSILE 732
1002 CTCTCATGTGGGAGAGCCAGTCCCTCACCATTAGACATGAGAGGCCATTC 1051
733 THRSEVALTHRASNLUSLYSTRLEUSERILEHISASPLEUGINASNI 749
1052 ACATCTGTAAACAACTGAAAACCTTGAGATATCATGACCTACAGAAATCA 1101
749 NARGLEUPROGLYLYLEUTHRASPSERLEUGLYASNLULYSANLEUT 766
1102 ACGGCTGCCGGGTGCTGTGACTGACAGCTTGGGTAACTGAAAACCTTA 1151
766 HRSYSEULIEMETASPSANILEYMETASNLUGLUSPALAIIELYS 782
1152 CNAAGCTCATATGATTAACATTAAGATGATGAACAGACATGCTTAATAAA 1201
783 LEUALAGLUGLYLEULYSANLEULYSLYSMETCYLSLEUPHEHISLEUTH 799
|||||

```

```

1202 CTAGCTGAAGGCTGTGAAAAACCTGAAGAAGATGTGTTATTCATTGTGAC 1251
799 RHISLEUSERASPIIEGLYGLULYMETASPTYRIIEVALYSELYS 816
1252 CCAGTGTCTGACATTGAGAGGGAATGATTAACATGATCAAGTCTGTGT 1301
816 ERSERGLUPROCYASPLEUGLULIIEGLINLEUVALSERCYSCYLEU 832
1302 CAAGTGAACCCCTGTGACCTTGAAGAAATTCATTAATCTCTCTGCTCTGG 1351
833 SERALASNALAYALLYSILEUVALAGINASNLUEUHSASNLUEVALY 849
1352 TCTGCAAAATGTGAGTGAATAATCTAGCTCAGAAATCTTCACAAATTTGGTCA 1401
849 SLEUSERILEUASPLEUSERGLYASNTYRIEUGLULYLSASPGLYASNG 866
1402 ACTGACCATTTCTGATTTATCAGAAAAATTACCTGAAAAAGAGGAAAG 1451
866 IUALLEUHHISGLULEUILEASPARGMETASNLVALLEUGLULINLEUTH 882
1452 AAGCTTCTCATGACTGATCGACAGCATGACGTGCTAGAACAGCTCAC 1501
883 ALALEUMETLEUPROTPRLCYASPPVALGLINGLYSERLEUSERSERIE 899
1502 GCACGTATGTCGCCCTGGGGCTGTGACGTGCAAGGACGAGCCTGAGCCCT 1551
899 ULEULYSHISLEUGLULYALPROGLINLEUVALYLSLEUGLYLEULYA 916
1552 GTTGAACAATTGGAGAGGTCCACACACTCGCAAGCTTGGGTTGAAAA 1601
916 SNTPRARGLEUTHRASPTRHGLUIIEARGIIELEUGLYALAPHEPHEGLY 932
1602 ACTGGAGACTCACAGATACAGAGATTAGAAATTTTACGTGATTTTGGTA 1651
933 LYSASNPROLEULYSANPHEGLINGLINLEUASNLUEUALAGLYASNARYA 949
1652 AAGAACCCTCTGAAAAAATTCAGCAGTTGAATTTGGCGGAAATGCTGT 1701
949 ISESERASPGLYTRPLEUALAPHEMETGLYVALPHEGLUASNLUEULYG 966
1702 GAGCAGTGTATGATGGCTTCTCATGGGTGATTTGAGAAATCTTAAC 1751
966 INLEUVALPHEPHEASPHESETHRLYSGLUPHELEUPROASPPROLA 982
1752 AATTACTGTTTTTGTACTTTAGACTTAAGAAATTTCTACTGTATCCAGCA 1801
983 LEUVALARGLYSLEUSERGLINVALLEUSERLYSLEUTHRPHLEUGINLI 999
1802 TTAGTCAGAAAACCTTAGCCAAAGTTATCCAAAGTTAACTTCTGTGCAAGA 1851
999 UALIAARGLEUVALGLYTRPGLINPHEASPSASPSASPLEUSERVALIIE 1016
1852 AGCTAGGCTTGTGGGTGGCAATTTGATGATGATGATGATGATGATGATTA 1901
1016 HRCGLYALAPHELYSLEUVALIATHRALA 1024
1902 CAGGTCTTTTAAACTAGTAAGTGTCT 1927

seq_name: /SID52/gcgdata/geneseq/geneseq/NA2001.DAT:AAH34171
seq_documentation_block:
ID AAH34171 standard; cDNA: 2735 BP.
XX
XX AAH34171;
XX
XX 03-SEP-2001 (first entry)
XX
XX Human colon cancer antigen encoding cDNA SEQ ID NO:1253.
DE
XX Human: colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; ss.
XX
XX Homo sapiens.
XX

```

```

XX  WO200122920-A2.
XX  05-APR-2001.
XX  28-SEP-2000: 2000MO-US26524.
XX  29-SEP-1999: 99US-0157137.
XX  03-NOV-1999: 99US-0163280.
XX  (HUMA-) HUMAN GENOME SCI INC.
XX  Ruben SM, Barash SC, Birse CE, Rosen CA:
XX  MPI: 2001-235357/24.
XX  P-PsDB: AAG74766.
XX  Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX  useful for preventing, diagnosing and/or treating colorectal cancers -
XX  Claim 1: Page 3017: 9803pp: English.
XX  AAH37943 to AAH37195 and AAG77788 represent human colon
XX  cancer-associated nucleic acid molecules (N) and proteins (P), where
XX  the proteins are collectively known as colon cancer antigens. The colon
XX  cancer antigens have cytostatic activity and can be used in gene
XX  therapy and vaccine production. N and P may be used in the prevention,
XX  diagnosis and treatment of diseases associated with inappropriate P
XX  expression. For example, N and P may be used to treat disorders
XX  associated with decreased expression by rectifying mutations or deletions
XX  in a patient's genome that affect the activity of P by expressing
XX  inactive proteins or to supplement the patient's own production of P.
XX  Additionally, N may be used to produce the colon cancer-associated Ps,
XX  by inserting the nucleic acids into a host cell and culturing the cell
XX  to express the proteins. N and P can be used in the prevention, diagnosis
XX  and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
XX  and AAB77788 represent sequences used in the exemplification of the
XX  present invention.
XX  N.B. Pages 666 to 682 and page 7053 of the sequence listing were
XX  missing at time of publication, meaning no sequences are present for
XX  SEQ ID NO:1027 to 1052, 7921 and 7922.
XX  Sequence 2735 BP; 791 A; 555 C; 512 G; 876 T; 1 other:

alignment_scores:
    Quality: 921.00      Length: 180
    Ratio: 5.117        Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AAH34171/rev ..

Align seg 1/1 to reverse of: AAH34171 from: 1 to: 2735
845 HisaaleuValLysLeuSerLeuAspLeuSerGluAsnTyrLeuGI 861
2727 CACAATTTGGTCAAACTGACATCTTGTATTCAGAAATTAACCTGGA 2678
861 ulysAspGlysnGluAlaLeuHisGluLeuIleAspArgMetAsnValL 878
2677 AAAAGATGGAAATGAAAGCTCTTCATGAACTGATCGACAGGATGAACGTGC 2628
878 euGluGlnLeuThrAlaLeuMetLeuProTfPGlyCYAspValGlnGly 894
2627 TAGAACAGCTACCGCACTGATGCTGCGCTGGGGCTGTGACGTGCAAGCC 2578
895 SerLeuSerSerLeuLeuLysHisLeuGlnLysValProGlnLeuVally 911
2577 AGCTGAGACACTGTGTGAACATTTGGAGAGAGGTCCCAACACTGCTCAA 2528
911 sleuGlyLeuLysAsnTfPAArgLeuThrAspThrGluLeuArgIleLeuG 928

```

```

2527 GCTTGGGTTGAAAAAAGTGGAGACTCAGATACAGATTAGAAATTTAG 2478
928 lYAlaphepGlyLysAsnProLeuLysAsnPheGlnGlnLeuAsnLeu 944
2477 GTGATTTTGTGGAGAGAACCTCTGAAAAACCTTCACACAGTGTGAATTTG 2428
945 AlaGlyAsnArgValSerSerAspGlyTfRpleuAlaPheMetGlyValPh 961
2427 GCGGAAATCGTGAGCAGATGATGATGCTTGCTTCATGGGGTGTAT 2378
961 eGluAsnLeuLysGlnLeuValPhePheAspPheSerThrLysGlnPheL 978
2377 TGAGAAATCTTAGCAATATGATGTTTTTGAAGTTAGTAAAGAAATTTTC 2328
978 euProAspProAlaLeuValArgLysLeuSerGlnValLeuSerLysLeu 994
2327 TACCTGATCCAGCATTTAGTCAGAAAACCTTAGCCAAAGTGTATCCAGTTA 2278
995 ThrPheLeuGlnGlnLysAlaArgLeuValGlyTfRpleuPheAspAspAs 1011
2277 ACTTTTCGCAAGAGCTAGGCTTGTGTGGCAATTTGATGATGATGA 2228
1011 pleuSerValIleThrGlyAlaPheLysLeuValThrAla 1024
2227 TCTCAGTGTATTACAGGTGCTTTTAAACTAGTAAGTGTCT 2188

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT.AAX58000
seq_documentation_block:
ID AAX58000 standard: cDNA to mRNA; 5984 BP.
XX AAX58000:
XX 20-JUL-1999 (first entry)
XX
XX Gonadotropic hormone coding sequence.
XX
XX Gonadotropic hormone; excessive ovulation animal; transgenic animal;
XX totipotent cell; somatic cell chromosome; ds.
XX Homo sapiens.
XX JP11113444-A.
XX 27-APR-1999.
XX
XX 14-OCT-1997: 97JP-0280830.
XX
XX 14-OCT-1997: 97JP-0280830.
XX
XX (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
XX (SAKA/) SAKAI H.
XX
XX MPI: 1999-320709/27.
XX P-PsDB: AAY14079.
XX
XX An excessive ovulation animal - useful for improving the
XX productivity of animals
XX
XX Claim 1: Page 7-9; 18pp; Japanese.
XX
XX This sequence encodes a gonadotropic hormone.
XX The invention relates to an excessive ovulation animal, which is a
XX transgenic animal with a totipotent cell containing a DNA fragment
XX containing a promoter sequence and a gonadotropic hormone coding
XX sequence. The DNA fragment is in the somatic cell chromosome. The
XX excessive ovulation animal is useful for improving the productivity of
XX animals. The method can improve the productivity of a useful animal.
XX
XX Sequence 5984 BP; 1777 A; 1267 C; 1267 G; 1673 T; 0 other:

alignment_scores:

```

Quality: 653.00 Length: 1108
Ratio: 1.074 Gaps: 42
Percent Similarity: 54.874 Percent Identity: 23.736

alignment_block:

US-09-697-089-2 x AAX58000 ..

Align seg 1/1 to: AAX58000 from: 1 to: 5984

```

69 PheLeuLysSerLeuLysGluTrpAsnTyrProLeuPheGlnAspLeuAs 85
   |||||:.....:|||||:.....:  |||||:
1321 TTCTCCAAATATGAAAGTCC...TCGCGGAAGTGACTCCAGACCTTCA 1367
   :ngLgln.....SerLeuPheHisGlnThrSergLysAspL 98
   :|||||:.....:|||||:.....:  |||||:
1368 GACCCGCGGGAAGTCTTGATTAATTAAGCAACAGTAAGCAATC 1417
   :.....:.....:.....:.....:  :
98 euAspAsp.....:.....:.....:.....: 100
   ||:||||
1418 TTGAGATTCATATAGCAGTTGTCTATAGTCCAGAAATGCACAGGGT 1467
   :.....:.....:.....:.....:  LeuAlaGlnAspLeuAsp 107
101 .....:.....:.....:.....: 117
1468 GAAGCCAGTGTCTTCAAGAGCAAAAGATGATGACGCTGAGAGC 1517
   || |||:|||||:.....:.....:  |||
107 PLeuTyrHisThrProSerPheLeuAspPhe.....:.....: 117
   || |||:|||||:.....:.....:  :
1518 AGCTTATACAGAGCCAGTTTCCGCCACATGTCTTCCCTGATATCTCT 1567
   :.....:.....:.....:.....:  :
118 .....:.....:.....:.....: 128
   :|||:|||||:.....:.....:  |||
1568 CCGATGTGGCCAGCAGACCTGTGGGCTGTGATCTGTCTATT..... 1611
   :.....:.....:.....:.....:  :
129 AsnLeuLysSerThrPheThrGluProValLeuTrpArgLysAspGlnH 145
   :.....:.....:.....:.....:  :
1612 ...GCTTCAAAACACATCAGCAAACTGTG.....:.....: 1638
   :.....:.....:.....:.....:  :
145 SHISHisArgValGluGlnLeuThrLeuAsnGlyLeuGlnAlaLeuG 162
   ||| ||||:|||||:.....:.....:  |||
1639 .....:.....:.....:.....: 1678
   :|||:|||||:.....:.....:  :
162 InSerProCysIleIleGluGlyLysSerGlyLysGlyLysSerThrLeu 178
   :|||:|||||:.....:.....:  :
1679 ACCTGTGCATGTGTGAGGAGTGAAAGTGAAGTGAAGACGAGCTCTC 1728
   :|||:|||||:.....:.....:  |||
179 LeuGlnArgTrpLeuIleMetLeuTrpGlySerGlyLysCysLysAlaLeuTh 195
   :|||:|||||:.....:.....:  |||
1729 CTGAAGAAATACCTTTCTGTGTGGCATGTGATGCTGCTCCCTGTAA 1778
   :|||:|||||:.....:.....:  :
195 rLysPheLysPheValPhePheLeuArgLeuSer.....:.....: 210
   :|||:|||||:.....:.....:  |||
1779 CAGGTTCACAGCTGTTTCTACCTCTCCCTTACTTCCACACAGCAGAGC 1828
   :|||:|||||:.....:.....:  :
210 LysIleLeuPheGluThrLeuCysAspGlnLeuLeuAspIleProGlyThr 226
   :|||:|||||:.....:.....:  :
1829 AGGGGCTGGCAGTATCTGTGACACAGCTCCTAGAGAAAGAGATCTC 1878
   :|||:|||||:.....:.....:  :
227 IleArgLysGlnThrPheMetAlaMetLeuLeuLysLeuArgGlnArgVa 243
   :|||:|||||:.....:.....:  :
1879 GTTACTGAAATGTGTCATGAGGAACATTATTCACAGAGTTAAAGATCAGGT 1928
   :|||:|||||:.....:.....:  :
243 LeuPheLeuLeuAspGlyTyrAsnGlnPhe.....:.....: 258
   :|||:|||||:.....:.....:  :
1929 CTTATTCCTTTTATGATGACTACAAAGAAATATGTCAATCCCTCAA..... 1974
   :|||:|||||:.....:.....:  :
258 ySProGluIleGluAlaLeuIleLysGluAsnHisArgPheLysAsnMet 274
   :|||:|||||:.....:.....:  :
1975 ....:.....:.....:.....: 2019
   :|||:|||||:.....:.....:  :
275 ValIleValThrThrThrThrGluCysLeuArgHisIleArgGlnPheG 291
   :|||:|||||:.....:.....:  :
2020 CTAATTGCTGTCTCCGTACAAACAGGCGCAGGACATCCGCGGATACCT 2069

```

```

291 ValAlaLeuThrAlaGluValGlyAspMetThrGluAspSerAlaGlnAla 308
   |||||:.....:.....:.....:  :
2070 AGAGACCATCTAGACATCAAAACATTTCCCTTTTAAATACGTCTGTGA 2119
   :|||:|||||:.....:.....:  :
308 euIleArgGluValLeuIleLysGlnLeuAla.....:.....: 322
   :|||:|||||:.....:.....:  :
2120 TATTACGAGAGCTCTTTTACATATATATGACTCGTCCGCAAGTTTATG 2169
   :|||:|||||:.....:.....:  :
323 LeuGlnIleGlnLysSerArgCysLeuArgAsnLeuMetLysThrProLe 339
   :|||:|||||:.....:.....:  :
2170 GTTTACTTTTGAAAGAAACCAAAAGTTTGCAAGATACAGAAACCTCTCT 2219
   :|||:|||||:.....:.....:  :
339 uphValValIleThrCysAlaIleGlnMetGlyLysSergLysPheHis 356
   :|||:|||||:.....:.....:  :
2220 CTTTGTGGCGCGCATCTGTCTCATCTGTTTGTCAAGTATCTTTGACCCAT 2269
   :|||:|||||:.....:.....:  :
356 eRhIsthrGlnThrThrLeuPheHisThrPheTyrAspLeuLeuIleGln 372
   :|||:|||||:.....:.....:  :
2270 CCTTGATGATGTGGCGTGTTCACGTCTATATGGAACGCTTTCTCTTA 2319
   :|||:|||||:.....:.....:  :
373 LysAsnLysHisLysHisLysGlyValAlaAlaSerAspPheLeuArg 388
   :|||:|||||:.....:.....:  :
2320 AGGAACAAA.....:.....:.....:.....: 2351
   :|||:|||||:.....:.....:  :
389 SerLeuAspHisCysGlyAspLeuAlaLeuGlnGlyValPheSerHisL 405
   :|||:|||||:.....:.....:  :
2352 AACTGTGCTCCTCTGTGTAGCTGGCTGAGCAAGGTTTTTTCATGTT 2401
   :|||:|||||:.....:.....:  :
405 ySPhAspPheGluLeuGlnAspVal.....:.....: 419
   :|||:|||||:.....:.....:  :
2402 GCATTAGTTTATGATGATGATCTCGCAGAACAGGAGGTTGATGAAGAT 2451
   :|||:|||||:.....:.....:  :
420 ValLeuLeuThrThrGlyLeuLeuCysLysTyrThrAlaGlnArgPhe 436
   :|||:|||||:.....:.....:  :
2452 GAAGATCTACCACTGTGATGATGACCAAAATTTACGCCAGCTAAAG 2501
   :|||:|||||:.....:.....:  :
436 sProLysTyrLysPhePheHisLysSerPheGlnGluTyrThrAlaGly 453
   :|||:|||||:.....:.....:  :
2502 ACCATTCTACCGGTTTTTAAGTCTGCTCCCAAGAAATTTCTGGCGGGA 2551
   :|||:|||||:.....:.....:  :
453 rGArgLeuSerSerLeuLeuThrSerHisGluProGluGluValThrLys 469
   :|||:|||||:.....:.....:  :
2552 TGAGCGTGAATTAACCTCGATTCAGATAGCGAGACATCAAGATTTG 2601
   :|||:|||||:.....:.....:  :
470 GlyAsnGlyTyrLeuGlnLysMetValSerIleSerAspIleThrSerTh 486
   :|||:|||||:.....:.....:  :
2602 GAGCTGATATCTTGAACAAATCAACTCACCCATGATGACTGTAAAGCC 2651
   :|||:|||||:.....:.....:  :
486 rTyrSerSerLeuLeuArgTyrThrCysGlySerSerValGluAlaThr 503
   :|||:|||||:.....:.....:  :
2652 CTACAACTATTTTGAATGATGTC.....:.....: 2695
   :|||:|||||:.....:.....:  :
503 rGAla.....:.....:.....:.....: 516
   :|||:|||||:.....:.....:  :
2696 AAGCAGGCGCCAAATTTGTCTCATTTGTCTCATTTAGTGGAATAC... 2742
   :|||:|||||:.....:.....:  :
517 CysLeuLeuGlyLeuSerIleAlaLysArgProLeuTrpArgGlnLys 533
   :|||:|||||:.....:.....:  :
2743 .....:.....:.....:.....: 2750
   :|||:|||||:.....:.....:  :
533 rLeuGlnSerValLysAsn.....:.....:.....: 539
   :|||:|||||:.....:.....:  :
2751 ATTGGAGATATATCTGAAATATGATGACTACTTAAGCAGCAGCCAGAA 2800
   :|||:|||||:.....:.....:  :
540 ..ThrThrGlnGlnGluIleLeuLysAlaIle.....:.....: 551
   :|||:|||||:.....:.....:  :
2801 TTTCACGTGAGATGACAGTACTTAGGGGATTTGCAAAATTTGTCCACA 2850
   :|||:|||||:.....:.....:  :
552 AsnSerPheValGluCysGlyIleHisLeu.....:.....: 561
   :|||:|||||:.....:.....:  :
2851 GCCTTACTTTTCAATGCTTTCAGAACATTTTACTGCTTTCCCTGAAAC 2900
   :|||:|||||:.....:.....:  :
562 ...TyrGlnGlnSerThrSerLysSerAlaLeuSerGlnGlu 577

```

```

2901 TGCCTATCAAGCAACACT...GTTCGTCGTCCTTCATTTGTTTGC 2947
2902 laphheglnlglylserserlyrleasnsersglyasnleproasp 593
2903 laphheglnlglylserserlyrleasnsersglyasnleproasp 593
2904 AATTCCTTCAAGGAGACACTGACTTGGTGCGCTTAACCTA...CAG 2994
2905 Tyrlaupheasp..... 597
2906 TACCTTTTCAGCACACCAGAAAGCTTCCTATGTTGAGAGACATCCACTT 3044
2907 Phephestuhsleup 603
2908 CCCAATACGAGAAATATAGACATCCAGACACATTTTCAGTTCTCG 3094
2909 roasnrcys.....Alaseralaleuaspheileysleuaspheyr 617
2910 AATTCCTTTCAGCAATACAGGTCGCAACTATATGATCAGACTATGCT 3144
2911 Glygly.....Alametalasertrpilu..... 625
2912 TCTGCTTTGAACCTATGATGAATGAGGAGCAATTTAGCTGAAGAAAGA 3194
2913 Lysalalagluaspt 631
2914 GGATTAATGTAAGACTATATGATATGACGCGGAGGATCACCAGACC 3244
2915 hrnglyglylehismetgluialaiprogluithrtyr...ileproser 646
2916 TTAGACTGCGCTATGGAACCTTCTCCAAAGCAGTACAGATTC... 3291
2917 Argalavalserleuphepshsnttrpysgluipheargthrleucl 663
2918 .....TCTCTAGA 3299
2919 uvalthrleuargasphepserlylsleuasnlyslinspillerhryl 680
2920 AGTCGATGTGATGATATGATGTTGTAGGCCAGATATGCTTGAGATTC 3349
2921 euglylysilepbeserlathrserleuargleuglnleuysarg 696
2922 TATAGACAGCTTTCTCAGCTTCACAGCCATCGAATCTCATTTAAGCAC 3399
2923 Cysalaglyvalalaglyserleuvalleuvalleuvalleuvalleuval 712
2924 AGCAGAGCTTTATAGAAAGCATCCGCCAGCTCTTGAGCTGTCTAAGGC 3449
2925 Asniletyrserleuvalleuvalleuvalleuvalleuvalleuvalleuval 729
2926 CTCTGTCCCAAGTGTCTCCATTAAGCAAGTGGAACTCAGCGCAGAAC 3499
2927 luarghisilethrserlathrserleuvalleuvalleuvalleuvalleuval 744
2928 AGAAGCTGTCTCTCAGCTTCCTGCTTCCTGAATCTCTGAAGTCCAGG 3549
2929 Aspleuglnasnlglnargleuproglyglyleuthrleuvalleuvalleuval 759
2930 ACAATTCAGCTCAGCAAGCAACCAATCTTCTTAAT.....CTGATTAAGCTT 3593
2931 ucllyasnleuvalleuvalleuvalleuvalleuvalleuvalleuvalleuval 775
2932 CCTGTGCTGAAGAAAGCTGTCTGTGATCTGAGGCGCAATATTAATGTTT 3643
2933 ..... 775
2934 TTTCAGTATTCCTGAAGAAATTTCCAACTTCACCATATGAGAAATTA 3693
2935 .....Asnlglnaspalaileuvalleuvalleuvalleuvalleuvalleuval 787
2936 TTGATTCAAATTTTCTGAGTATGATCTTCACCAATGATTAATTAAT 3743
2937 ulysasnleuvalleuvalleuvalleuvalleuvalleuvalleuvalleuval 802
2938 .....Thrhlsleus 802

```

```

3744 TCAAAATTCCTCAAAACCTTCATGTTTTCATCTGAGTAACTTCCTTT 3793
3745 eraspilleglygluylmetaspyrilevalleuvalleuvalleuvalleuval 818
3746 CGGATTTGGGTCTCTCATGACTATGCTTTGTTCC..... 3828
3747 ProCys...Aspleuglnaspalaileuvalleuvalleuvalleuvalleuval 834
3748 TGTAAAGAACTCACAGAAATTAAGTTTCGGATTCATCTTTT... 3870
3749 asnvalleuvalleuvalleuvalleuvalleuvalleuvalleuvalleuval 851
3750 CAAGCCCTCCCATTTGTTGCC...AGTTGCCAAATTTTATTTCTCTGA 3916
3751 erileleuaspheup...sergluasnlyrleuvalleuvalleuvalleuvalleuval 866
3752 AGATATTTAACTTGAAGCCAGCAATTTCTGATGAGAAACATCAAGAA 3966
3753 AlaLeuHisgluileuvalleuvalleuvalleuvalleuvalleuvalleuval 883
3754 AAATTTGCTCAATTTTATGTTCTCTTATGATCACTGCAAGAA..... 4008
3755 aleuvalleuvalleuvalleuvalleuvalleuvalleuvalleuvalleuval 900
3756 TTTATCTCTCTCTACTGCGGATGGAATTTATGAGTGGCCAACTGATCA 4057
3757 eulysnleuvalleuvalleuvalleuvalleuvalleuvalleuvalleuvalleuval 916
3758 TCCAGCACTGTCAGAGCTTATGTTCTGCGAGTCTCTCATTTTTCAG 4107
3759 TrpArgleuthrleuvalleuvalleuvalleuvalleuvalleuvalleuvalleuval 928
3760 ACTTTGAATGATGACAGCGTGTGGAATTTGCCAAAGTGAAGCAATCAGTGG 4157
3761 yAlaphepheglnlylsasnproleuvalleuvalleuvalleuvalleuvalleuval 945
3762 AGCTTTC.....CAGAACTTGAGCAACCTTAAGCTTT 4189
3763 laglyasn...Argvalserlathrserleuvalleuvalleuvalleuvalleuvalleuval 960
3764 CAATCAATCAAGATTTACAGAGAGATACAGAAATTTCTTTCAGACA 4239
3765 Phegluasnleuvalleuvalleuvalleuvalleuvalleuvalleuvalleuvalleuval 974
3766 CTGCAACATGACCAAACTTCAGAGATTTGACATCTCCAGCATTTTCAC 4289
3767 rlysglnleuvalleuvalleuvalleuvalleuvalleuvalleuvalleuvalleuval 991
3768 AGAGTGTATCAAGCTCAGGCCACACAGCTCAAGTCTTGTGATCAATGTG 4339
3769 euserlylsleuthrleuvalleuvalleuvalleuvalleuvalleuvalleuvalleuval 1007
3770 TGTATGACATCAAGGCTCATTTAGACTGACATGATTAAGTTGAGCTCTTG 4389
3771 Aspaspasleuvalleuvalleuvalleuvalleuvalleuvalleuvalleuvalleuval 1015
3772 GATGCAAGATGATTTGATTTGCTT 4413
3773 seq_name: /stids2/gcgdata/geneseq/geneseqn/NA1999.DAT.AAX56272
3774 seq_documentation_block:
3775 ID AAX56272 standard; cDNA to mRNA; 5984 BP.
3776 AC AAX56272;
3777 XX
3778 DT 20-JUL-1999 (first entry)
3779 DE Human apoptosis inhibiting protein encoding cDNA #1.
3780 XX Human; apoptosis inhibitory protein; apoptotic disease; diagnosis;
3781 KW spinal muscular atrophy; ds.
3782 OS Homo sapiens.

```

XX	JP11116599-A.
PN	
XX	27-APR-1999.
PD	
XX	14-OCT-1997; 97JP-0280831.
PE	
XX	14-OCT-1997; 97JP-0280831.
PR	
XX	
XX	
PA	(KAGA-) KAGAKU GIYUTSU SHINKO JIGYODAN.
XX	
XX	WPI; 1999-323531/27.
DR	P-PSDB; AAY09539.
XX	
PT	New apoptosis inhibitory protein - useful for determining mechanism of various apoptotic diseases e.g. human spinal muscular atrophy
XX	
PS	Claim 7; Page 11-13; 16pp; Japanese.
XX	
XX	The present sequence encodes a human apoptosis inhibitory protein.
CC	The apoptosis inhibitory protein is useful for the elucidation of the mechanism of various apoptosis diseases such as human spinal muscular atrophy and the diagnosis, the prevention and the treatment of such diseases.
CC	
XX	
XX	

Sequence 5384 BP; 1777 A; 1267 C; 1267 G; 1673 T; 0 other;


```

3744 TCAAAATTCACAACTTCATGTTTCCATGCACTGTAAGTCACTTCTTT 3793
802 eaisprielegyluglymetasptryrilevallysserleuserSerlu 818
3794 CGGATTTTGGGTCTCTCATGACTATGCTGTGTTCC..... 3828
819 Procys...AspleuglugliugliegluleuValSerCysCysleuseral 834
3829 ...TGTAGAACTCACAGAAATTAAGTTTCGATTCATTTT..... 3870
834 aasnAlaVallyslleuAlaaglInasnleuHlssnleuVallyslus 851
3871 .CAAGCGCTCCATTTGTTCC...AGTTGCCAAATTTATTTCCTCGA 3916
851 erlleuasnpleu...SerGlusnTyrlleuGlulysaspGlyAsnGlu 866
3917 AGATATTAAATCTTGAAGGCCACCAATTTCTGTATGAGGAACATCAGAA 3966
867 AlaLeuHlsgluLeuIleasparGmetasnValleugluInleuThrAl 883
3967 AAATTTGCCACATTTTACGTTCTCTTAGTAACCTGGAGAA..... 4008
883 aleuMetleuProTrrpGlyCysaspValInglySerleuserSerleu 900
4009 .TTGATCCCTTCCTAGCGGATGGAATTTATCGAGTGGCCAACTGATCA 4057
900 eulysHlslleugluGluValPrrglInleuVallyslleuglyLysAsn 916
4058 TCCAGCGTGTCCAGACCTTCATTTGCTCCGAGTCTCTCATTTTTCAG 4107
917 TrpArgleuThrAspThr.....GluIleArglleuGlu 928
4108 ACTTTGAATATATACACCGTGTGGAATTTCCAAAGTATACCAATCAGTG 4157
928 yAlaPhePheGlyLysAsnProleuLysAsnPheGlnInleuAsnleuA 945
4158 AGGTTTC.....CAGAAACTTGAGAACCTTAAAGCTTT 4189
945 laGlyAsn...ArgValSerSerAspGlyTrpPheAlaPheMetGlyVal 960
4190 CAATTCATATCAAGATTACAGAGGAGATACAGAAATTTCTTTCAGACA 4239
961 PheGluAsnleuLysGlnleuValPhePheAspPheSer.....Th 974
4240 CTGGACACACATGCCAACTTGCAGAGTTGAGACATCTCCAGCATTTTCA 4289
974 rLysGluPheLeuProAspPrrAlaLeuValArgLysLeuSerGlnVal 991
4290 AGAGTGTATCAAGCTCAGGCCACACAGCAAGTCTTTGAGTCAATGTG 4339
991 euserLysleuThrPheleugluAlaArgleuValGlyTrpGlnPhe 1007
4340 TGTTAGCAGATCAAGGCTCTTAGTGAACATGTTAAGTGGCTCTTG 4389
1008 AspAspAspAspLeuSerValIle 1015
4390 GATGCACATCATATTCATTCCTT 4413
seq_name: /SIDS2/gc9data/geneseq/geneseqn/NA1997.DAT: AAT71265
seq_documentation_block:
ID AAT71265 standard; cDNA: 6124 BP.
XX
XX AAT71265;
XX
XX 06-OCT-1997 (first entry)
XX
XX Neutonal apoptosis inhibitor protein (NAIP) cDNA.
XX
XX Neutonal apoptosis inhibitor protein; NAIP; diagnosis;
XX gene therapy; cancer; AIDS; amyotrophic lateral sclerosis;
XX spinal muscular atrophy; ds.

```

```

XX Homo sapiens.
OS
XX Key Location/Qualifiers
FH CDS 292..4503
FT CDS /*tag= a
XX
XX W09726331-A2.
XX
XX 24-JUL-1997.
XX
XX 17-JAN-1997; 97WO-IB00142.
XX
XX 19-JAN-1996; 96GB-0001108.
XX
XX (UNOT-) UNIV OTTAWA.
XX
XX Korneluk RG, Mackenzie AE, Robertson G, Roy N, Tamai K;
XX WPI: 1997-385335/35.
XX P-PSDB; AAM20032.
XX
XX New neuronal inhibitor of apoptosis - useful for diagnosing and
XX treating, e.g. cancer, AIDS or amyotrophic lateral sclerosis
XX
XX Claim 28; Fig 6A-I; 102pp; English.
XX
XX A cDNA clone (AAT71265) codes for novel human neuronal apoptosis
XX inhibitor protein (NAIP) (AAM20032). It was isolated from a human
XX foetal spinal cord cDNA library by probing with the genomic insert
XX in cosmid 25086, containing a CAT locus, and completion of gaps
XX in the sequence by further library screenings. The NAIP gene, on
XX chromosome 5q13.1, is about 56 kb long and has 18 exons (see also
XX AAT71266). NAIP is a negative regulator of apoptosis, partic.
XX neuronal apoptosis and, when deficient or absent, contributes to
XX neurodegenerative phenotypes such as spinal muscular atrophy (SMA)
XX and amyotrophic lateral sclerosis. NAIP nucleic acids are useful
XX for protein expression in host-vector systems, as probes/primers,
XX and in gene therapy to inhibit apoptosis for treatment of AIDS,
XX neurodegenerative disease, myelodysplastic syndromes or ischaemic
XX injury, or (antisense) to increase apoptosis.
XX
XX Sequence 6124 BP; 1811 A; 1297 C; 1286 G; 1730 T; 0 other;
SQ
alignment_scores:
Quality: 653.00 Length: 1108
Ratio: 1.074 Gaps: 42
Percent Similarity: 54.874 Percent Identity: 23.736
alignment_block:
US-09-697-089-2 x AAT71265 ..
Align seg 1/1 to: AAT71265 from: 1 to: 6124
69 PheLeuLysSerLeuLysGluTrpAsnTyrrProLeuPheGlnAspLeuAs 85
1321 TTTTCTCAAAATATGAAAGTCC...TCTCGGAAAGTACTCCACACCTTCA 1367
85 nGlyGln.....SerLeuPheHlsglnThrSerGluGlyAspL 98
1368 GAGCCGTGTGGAACCTTGTGAATTAAGTCAACCAACCAAGCAAGCAATC 1417
98 euaSpAsp..... 100
1418 TTGAGATTCATATGACGTGTGCTCTATGATGCCAGAAATGGACAGGCT 1467
101 .....LeuAlaGlnAspLeuLysAs 107
1468 GAAGCCAGTGTTCACAGAGCAAGCAATCTGAATGACGAGCTGAGAC 1517
107 PleuTyrrHlsthPrSerPheLeuAsnPhe..... 117

```

1518 AGCTTATACCAAGCCAGTTCTCCGCCACATGCTTGTGATATCTCTT 1567
118TyrProLeuGluGluAspIleAspIleIlePhe 128
1568 CGCATCTGCCACGACGACCTTGTGGGCTGTGATCTGTCTATT..... 1611
129 AsnLeuLysSerThrPheThrGluProValLeuTrpArgLysAspGlnI 145
1612 ...GCTTCAAAACATCAGCAAACTGTG..... 1638
145 ShHisIAsrValGluGlnLeuThrLeuAsnGlyLeuLeuGlnAlaLeu 162
1639CAAGAACTCTGTGTCTGTCTGAGAGTCTTTGGCAACTTGA 1678
162 InSerProCysIleIleGluGluGluSerGlyLysGlyLysSerThrLeu 178
1679 ACTGTCTCATGTGTGTGGAGGTGAGTGAAGTGAAGAAAGCGTCTCTC 1728
179 LeuGlnArgIleAlaMetLeuTrpGlySerGlyLysLysAlaLeuThr 195
1729 CTGAAGAAATAGCTTTCTGTGGGATCTGTGATGCTGTCCCTGTAA 1778
195 rLysPheLysPheValPhePheLeuArgLeuSer.....ArgAlaGln 210
1779 CAGGTTCACAGCTGTTTCTACTCTCCCTTAGTCCACAGACACAGACG 1828
210 lGlyLeuPheGluThrLeuCysAspGlnLeuLeuAspIleProGlyThr 226
1829 AGGGCTGGCCAGTATCATCTGTGACCAAGCTCTGAGAAAGACGATCT 1878
227 lLeuArgLysGlnThrPheMetAlaMetLeuLeuLysLeuArgGlnArg 243
1879 GTTACTGAATGTGCATAGAGAACTATTCAGACAGTTAAAGAACTAGCT 1928
243 lLeuPheLeuLeuAspGlyTyrAsnGluPhe.....LysProGlnAsn 258
1929 CTATTCTCTTATAGTACTACAAAGAAATATGTCATCTCTCA..... 1974
258 ySProlGluIleGluAlaLeuIleLysGluAsnHisArgPheLysAsnMet 274
1975GTCATAGAAACTGATTTCAAAAAACCACTTATCCCGACCTGC 2019
275 ValIleValIleThrThrThrGluCysLeuArgHisIleArgGlnPheG 291
2020 CTATTGATGTGCTCGTACAAACAGGGCCAGGACATCCGCCGATACCT 2069
291 yAlaLeuThrAlaGluValAlaGlyAspMetThrGluAspSerAlaGlnAla 308
2070 AGAGACATTTCTAGAGATCAAGCATTTCCCTTATATATACTGTCTGA 2119
308 euIleArgGluValLeuIleLysGluLeuAla.....GluGlyLeuLeu 322
2120 TATTCAGGAAGCTCTTTACATAATATGACTGTCTGCAAAAGTTTATG 2169
323 LeuGlnIleGlnLysSerArgCysLeuArgAsnLeuMetLysThrPro 339
2170 GTTATCTTGGAAAGAACCAAGTTTGCAGAGATACAGAAAACTCTCT 2219
339 uPheValValIleThrCysAlaIleGlnMetGlyLysSerGluPheHis 356
2220 CTTTGTGGCGGAGATCTGTGCTCATGTGTTTCAGATATCTTTGACCAT 2269
356 eHisThrGlnThrThrLeuPheHisThrPheTyrAspLeuLeuIleGln 372
2270 CTTTGTGATGTGGCTGTTTTCAGATCTTATGAAAGCGCTTTCTCTTA 2319
373 LysAsnLysHisLysHisLysGlyValAlaAlaAspAspPheIleArg.. 388
2320 AGGAACAAA.....GCGACAGCTGAATTTCTCAAGC 2351
389 ..SerLeuAspHisCysGlyAspLeuAlaLeuGluGlyValPheSerHis 405
2352 AACGTGTCTCTGTGTGAGCTGTGCTTGAAGGCTTTTTCATGT 2401

405 ySPheAspPheGluLeuGlnAspVal.....SerSerValAsnGluAsp 419
2402 GCTTGAATTTAATGATGATGATCTCGCAGAAAGCGGGGTGATGAACAT 2451
420 ValLeuLeuThrThrGlyLeuLeuCysLysTyrThrAlaGlnArgPhe 436
2452 GAAGATCTAACCTGTGCTGTATGAGCAAAATTTACAGCCAGAGACTAAG 2501
436 sProlLysThrLysPhePheHisLysSerPheGlnGluThrAlaGly 453
2502 ACCATTCTACCGGTTTAACTCTGCTCCCAAAATTTCTTGGGGGA 2551
453 rArgLeuSerSerLeuLeuThrSerHisGluProGlnValValThrLys 469
2552 TGAGCGCTATTGAACCTCTGATTCAGATAGCAGAAACATCAAGATTG 2601
470 GlyAsnGlyTyrLeuGlnLysMetValSerIleSerAspIleThrSer 486
2602 GGACTGTATCATTTGAACAAATCAACTACCCATGATGACTGTGAAGGC 2651
486 rTyrSerSerLeuLeuArgTyrThrCysGlySerSerValGluAlaThr 503
2652 CTACACAAATTTTGAACATATGTC.....TCCAGCTCCCTTCACAA 2695
503 rGla.....ValMetLysHisLeuAlaAlaValTyrGlnHisGly 516
2696 AAGCAGGCCCAAAATGTGTCTCATTTGCTCCATTTAGTGATGATAC... 2742
517 CysLeuLeuGlyLeuSerIleAlaLysArgProLeuTrpArgGlnLys 533
2743AAAGATC 2750
533 rLeuGlnSerValLysAsn..... 539
2751 ATTGAGAAATATATCTGAATATGATGACTACTTAAGCACACAGCAAA 2800
540 ..ThrThrGluGlnGluIleLeuLysAlaIle.....AsnIle 551
2801 TTTCACCTCAGATGCAGTTACTAGGGGATGTGCAAAATTTGTCCACA 2850
552 AsnSerPheValGluCysGlyIleHisLeu..... 561
2851 GCTTACTTTTAAAGTTTTCAGAAACATTTACTGTCTTCCCTGAAAC 2900
562TyrGlnLysSerThrSerLysSerAlaLeuSerGlnLupheGlu 577
2901 TGCTTATCAAGCAACT...GTTCGTGCGTGTCTCTCATTTGTTTGC 2947
577 lAphePheGlnGlyLysSerLeuTyrIleAsnSerGlyAsnIleProAsp 593
2948 AATTCCTTCAAGGGAGAACACTGACTTTGGGTGGCTTAACTTA...CAG 2994
594 TyrLeuPheAsp..... 597
2995 TACTTTTTCAGACCCAGAAAGCTTTCATTTGAGGACATCCACTT 3044
598PhePheGlnHisLeuP 603
3045 CCAATACAGAGAAATAGACATCACCCAGAGACATTTTTCAGTTCTGG 3094
603 rAsnCys.....AlaSerAlaLeuAspPheIleLysLeuAspPheTyr 617
3095 AAACATGTTTTCAGAAATCAGACAGTCCAACTATAGATCAGGACTATGCT 3144
618 GlyGly.....AlaMetAlaSerTrpGlu..... 625
3145 TCTGCTTTGAACCTATGATGAATGGAGGCAAAATTTAGCTGAAGAAAGA 3194
626LysAlaIleGluAsp 631
3195 GGATATGTAAAGACTATATGATATGACGACGACGACATCACACAGAC 3244

1008 AspaspaspLeuSerValIle 1015
 ||| |||||
 4399 GATCAGATGATATGATGCTT 4422

seq_name: /STD2/gcgdata/geneseq/geneseq/NA1997.DAT:AA71266

seq_documentation_block:

ID AA71266 standard; cDNA; 6228 BP.

XX AC AA71266;

XX DT 06-OCT-1997 (first entry)

XX Neutroal apoptosis inhibitor protein (NAIP) cDNA.

XX Neutroal apoptosis inhibitor protein; NAIP: diagnosis;

KW gene therapy; cancer; AIDS; amyotrophic lateral sclerosis;

XX spinal muscular atrophy; ds.

XX OS Homo sapiens.

XX Key

XX Location/Qualifiers

XX CDS

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

/*tag= q
 /number= 15
 4078..4242
 /*tag= r
 /number= 16
 4243..6228
 /*tag= s
 /number= 17

MO9726331-A2.

24-JUL-1997.

17-JAN-1997: 97MO-IB00142.

19-JAN-1996: 96GB-0001108.

(UYOT-) UNIV OTTAWA.

Korneluk RG, Mackenzie AE, Robertson G, Roy N, Tamai K;

WPI: 1997-385335/35.

P-P-SDB; AAM20033.

New neuronal inhibitor of apoptosis - useful for diagnosing and

treating, e.g. cancer, AIDS or amyotrophic lateral sclerosis

Example 4; Fig 7A-L; 102pp; English.

A cDNA clone (AA71266) codes for novel human neuronal apoptosis
 inhibitor protein (NAIP) (AAM20033). It was isolated from a human
 foetal spinal cord cDNA library by probing with the genomic insert
 in cosmid 25086, containing a CAT locus, and completion of gaps
 in the sequence by further library screenings. The NAIP gene, on
 chromosome 5q13.1, is about 56 kb long. NAIP is a negative
 regulator of apoptosis, partic. neuronal apoptosis and, when
 deficient or absent, contributes to neurodegenerative phenotypes
 such as spinal muscular atrophy (SMA) and amyotrophic lateral
 sclerosis. NAIP nucleic acids (see also AA71264 and AA71265) are
 useful for protein expression in host-vector systems, as probes or
 CC primers, and in gene therapy to inhibit apoptosis for treatment of
 CC AIDS, neurodegenerative disease, myelodysplastic syndromes or
 CC ischaemic injury, or (antisense) to increase apoptosis.

Sequence 6228 BP; 1826 A; 1330 C; 1304 G; 1767 T; 1 other;

alignment_scores:

Quality: 652.00 Length: 1103

Ratio: 1.069 Gaps: 41

Percent Similarity: 55.304 Percent Identity: 23.663

alignment_block:

us-09-697-089-2 x AA71266 ..

Align seg 1/1 to: AA71266 from: 1 to: 6228

69 PheLeuYsSerLeuYsGLuTPraSnTYrProLeuPheGlnAspLeuAs 85

1425 TTTCTCCAAATATGAGAGCTCC...TTCGCGAGTACTCCAGACTTCA 1471

85 nGlyGln.....SerLeuPheHisGlnThrSerGluLysAspL 98

1472 GAGCGGTGAGTACTTGTGAATTACTGGAACCAACCAAGTAAAGCAATC 1521

98 euAspAsp..... 100

1522 TTGAAGATTCAATAGACAGTGTGCTCATATAGTCCAGAAATGACACAGGCT 1571

101LeuAlaGlnAspLeuLysAs 107

1572 GAAGCCAGTGTTCATCAAGAGCAAAAGATCTGAATGAGCAGCTGAGAGC 1621


```

3299 GGATAATGTAAGACTATATGATATGCAGCGAGGGCATCCACGACC 3348
631 hrglyglyIleHisMetGluGluIuaIaProGluThrYr...IleProSer 646
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3349 TTAGTACTGGGTATTGGAACTTCTCCAAAGCAAGTACAGATTCCC... 3395
647 ArgAlaValSerLeuPhePheAsnTrpLysGlnIupheArgThrLeuG 663
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3396 .....TGTCTAGA 3403
663 uValThrLeuArgAspPheSerLysLeuAsnLysGlnAspIleThrYrL 680
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3404 AGTCAGTGTGATGATATGATGTTGAGCCAGAGATATGCTTGAGATT 3453
680 euGlyLysIlePheSerSerAlaThrSerLeuArgLeuGlnIleLysArg 696
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3454 TAAATGACAGTTTCTCAAGTTCACAGCGCATCGAACTCATTTAAACAC 3503
697 CysAlaGlyValAlaGlySerLeuSerLeuValLeuSerThrCysLys.. 712
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3504 AGCAGAGGCTTTATAGAAAGCATCCGCCACGCTTGAGCTGTCTAAGGC 3553
713 AsnIleTyrSerLeuMetValGluAlaSerProLeuThrIleGluAspG 729
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3554 CTCTGTACCAAGTGTCTCCATAGCAAGTTGGAACTCAGCGACGCCGAC 3603
729 IuaArgHisIleThrSerValThrAsnLeuLysThrLeuSerIleHis... 744
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3604 AGAAAGTGTCTTCACCCCTGCTCCCTCGGAATCTCTTGAAGTCTCAGGG 3653
745 .....AspLeuGlnAsnGlnArgLeuProGlyLysIleThrAspSerLe 759
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3654 ACAATCCAGTCCACAGACCAAAATCTTCTCTAAT.....CTGATTAAGTT 3697
759 uGlyAsnLeuLysAsnLeuThrLysLeuIleMetAspAsnIleLysMet. 775
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3698 CCTGTGCTGTAAGAACTGTCTGTGATCTGGAGGCGCAATATTAATGTTT 3747
775 ..... 775
3748 TTTCAGTCATTCCTGAAAGAAATTTCCAAACTTCACCATATGAGAAATTA 3797
776 .....AsnGluGluAspAlaIleLysLeuAlaGluGlyLe 787
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3798 TTGATCCAAATTTAGCTGATATGATATCTCCAAACTAGTAAATTAAT 3847
787 uLysAsnLeuLysLysMetCysLeuPheHisLeu.....ThrHisLeus 802
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3848 TCAAAATTTCCAAACCTTCATGTGTTTCCATCTGAAGTAACTTCTTTT 3897
802 eArpIleGlyGluGlyMetAspTrpIleValLysSerLeuSerSerGlu 818
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3898 CGGATTTTGGGTCTCATGACTATGCTTGTTC..... 3932
819 ProCys...AspLeuGluGluIleGlnLeuValSerCysCysLeuSerAl 834
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3933 ...TGTAAGAAACACACAAATTAAGTTTCGGATTCATTTT... 3974
834 aAsnAlaValLysIleLeuAlaGlnAsnLeuHisAsnLeuValLysLeus 851
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3975 .CAAGCCGTCCTCATTTGTTGCC..AGTTGCCAAATTTATTTCTCTGA 4020
851 eIleLeuAspLeu...SerGluAsnTrpLeuGluLysAspGlyAsnGlu 866
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4021 AGATAATTAATCTTGAAGGCCAGCAATTTCTGATGAGGAACAATCAGAA 4070
867 AlaLeuHisGluLeuIleAspArgMetAsnValLeuGluIleuThrAl 883
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4071 AAATTTCCCTACATTTAGTCTTCTTAACTGAACTCGAAGAA..... 4112
889 aLeuMetLeuProTrpGlyCysAspValGlnGlySerLeuSerSerLeu 900
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4113 .TTGATCTCTCTACTGTGGGATGGAATTTATGAGTGGCCAAACTGATCA 4161

```

```

900 euLysHisLeuGluGluIuaIaProGlnLeuValLysLeuGlyLeuLysAsn 916
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4162 TCCAGCACTGTCCAGACCTTCATGTCTCCGAGTCCCTCATTTTTCAG 4211
917 TrpArgLeuThrAspTrhGluIleArgIleLeuGlyAlaPhePheGlyLy 933
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4212 ..ACTTGAATGATGACACGCTGTGGAAATTCGCAAGTACGAAATCA 4258
933 sAsnProLeuLysAsnPheGlnGlnLeuAsnLeuAlaGlyAsn...ArgV 949
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4259 TGCAGGTTCCAGAAACTTGAGAACTTAACCTTTTCATCAATCAACAAG 4308
949 aIleSerAspGlyTrpLeuAlaPheMetGlyValPheGluAsnLeuLys 965
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4309 TTACAGAGGAAGATACAGAAATTTCTTCAAGCACTCGACAAACATGGCA 4358
966 GlnLeuValPhePheAspPheSer.....ThrLysGluPheLeuPr 979
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4359 AACTTGCAGAGTGTGACATCTCCAGGCATTTTCACAGAGTGTATCAAGC 4408
979 oAspProAlaLeuValArgLysLeuSerGlnValLeuSerLysLeuThrP 996
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4409 TCAGGCCACAAACAGTCAAGTCTTTGAGTCAATGTGTACGACTACCAA 4458
996 heLeuGlnGluAlaArgLeuValGlyTrpGlnPheAspAspAspLeu 1012
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4459 GGCTCATATAGACTGAAACATGTTAAGTTGGCTCTGTGATGACAGATATAT 4508
1013 SerValIle 1015
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4509 GCATTTGCTT 4517

seq_name: /STDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:AAx58001
seq_documentation_block:
ID AAX58001 standard; cDNA to mRNA; 5366 BP.
XX
AC AAX58001;
XX
DT 20-JUL-1999 (first entry)
XX
DE Gonadotropic hormone coding sequence.
XX
KW Gonadotropic hormone; excessive ovulation animal; transgenic animal;
KW totipotent cell; somatic cell chromosome; ds.
XX
OS Homo sapiens.
XX
PN JP11113444-A.
XX
PD 27-APR-1999.
XX
PF 14-OCT-1997; 97JP-0280830.
XX
PR 14-OCT-1997; 97JP-0280830.
XX
PA (KAGA-) KAGAKU GIUTUSU SHINKO JIGYODAN.
PA (SAKA/) SAKAI H.
XX
DR WPI; 1999-320709/27.
DR P-PDB; AAY14080.
XX
PT An excessive ovulation animal - useful for improving the
PT productivity of animals
XX
PS Claim 1; Page 9-11; 18pp; Japanese.
XX
CC This sequence encodes a gonadotropic hormone.
CC The invention relates to an excessive ovulation animal, which is a
CC transgenic animal with a totipotent cell containing a DNA fragment
CC containing a promoter sequence and a gonadotropic hormone coding
CC sequence. The DNA fragment is in the somatic cell chromosome. The

```

CC excessive ovulation animal is useful for improving the productivity of
 CC animals. The method can improve the productivity of a useful animal.
 XX

Sequence 5366 BP; 1483 A; 1173 C; 1222 G; 1488 T; 0 other;

Alignment_scores:

Quality: 593.50 Length: 1094
 Ratio: 1.023 Gaps: 42
 Percent Similarity: 53.016 Percent Identity: 23.675

Alignment_block:

US-09-697-089-2 x AAX58001 ..

Align seg 1/1 to: AAX58001 from: 1 to: 5366

```

69 PheLeuLysSerLeuLysGluThrPasnTrpProLeuPheGlnAspLeuS 85
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1321 TTTCTCCAAATATGAGTC...TCTGGGAGAGTCACTCCAGACCTTCA 1367
85 nGlyGln.....SerLeuPheHisGlnThrSerGluGlyAspL 98
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1368 GAGCGGTGGTGAATTTGTGAATTACTGGAAACACAGTGAAGCAATC 1417
98 euAspAsp..... 100
|||||:|||||:|||||:|||||:|||||:|||||:
1418 TTGAAGATTTCATAGCAGTTGGTCTATAGTCCAGAAATGCGACAGGT 1467
101 .....LeuAlaGlnAspLeuLysAs 107
|||||:|||||:|||||:|||||:|||||:|||||:
1468 GAAGCCAGTGGTTTCAAGAGGCAAGAAATCTGAATGAGCAGCTGAGAC 1517
107 PheUtyrHisThrProSerPheLeuAsnPhe..... 117
|||||:|||||:|||||:|||||:|||||:|||||:
1518 ACCTTATACAGCGCAGTTTCCGCACATGTCCTTGGTATATCTTT 1567
118 .....TyrProLeuGlyLysAspLeuAspLeuLephe 128
|||||:|||||:|||||:|||||:|||||:|||||:
1568 CCGATCTGGCCAGCGACCTGGGCTGTGATCTGTCTATTT..... 1611
128 AsnLeuLysSerThrPheThrGluProValLeuTyrPargLysAspGlnHi 145
|||||:|||||:|||||:|||||:|||||:|||||:
1612 ...GCTTCAAAACACATCAGCAAACTGTG..... 1638
145 sHisHisArgValGluGlnLeuThrLeuAsnGlyLeuGlnAlaLeuG 162
|||||:|||||:|||||:|||||:|||||:|||||:
1639 .....CAGAAACCTCTGTGCTGCTGAGGCTTTGGCAACTTGA 1678
162 InSerProCysIleIleGluGlyLysSerGlyLysGlyLysSerThrLeu 178
|||||:|||||:|||||:|||||:|||||:|||||:
1679 ACTGTGTATCTGTGGAGGCTGAAGCGAAGTGAAGAGCGGTCTTC 1728
179 LeuGlnAlaGlyIleAlaMetLeuThrPheLysGlyLysCysLysAlaLeuTh 195
|||||:|||||:|||||:|||||:|||||:|||||:
1729 CTGAAGAAATAGCTTCTTCTGTGGCATCTGATCTGCTCCCTGTAAA 1778
195 rLysPheLysPheValPhePheLeuArgLeuSer.....ArgAlaGln 210
|||||:|||||:|||||:|||||:|||||:|||||:
1779 CAGGTTCCAGCTGGTGTCTTCTACCTCTCCCTTAGTGTACACAGACAGAG 1828
210 LysGlyLeuPheGlnThrLeuCysAspGlnLeuLeuAspLleProGlyThr 226
|||||:|||||:|||||:|||||:|||||:|||||:
1829 AGGGCTGGCCAGTATCTGTGACAGCTCTCTGAGAAAGAAAGAGATCT 1878
227 IleArgLysGlnThrPheMetAlaMetLeuLeuLysLeuArgGlnArgVal 243
|||||:|||||:|||||:|||||:|||||:|||||:
1879 GTTACTGAATCTGCATGAGCAATCTATCCAGCAGTTAAAGAACAGGT 1928
243 lleuPheLeuLeuAspGlyTyrAsnGluPhe.....LysProGlnAsn 258
|||||:|||||:|||||:|||||:|||||:|||||:
1929 CTTATCTCTTTAGATGACTACAAAGAAATATGTTCAATCCCTCAA.... 1974
258 yspProGluIleGluAlaLeuIleLysGluAsnHisArgPheLysAsnMet 274
```

```

1975 .....GTATAGAAACATGATTCAAAAAAACCACTTATCCCGAGCTCC 2019
275 ValIleValThrThrThrGluCysLeuArgHisIleArgGlnPheG 291
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2020 CTATTGATGTCTGCTCGGTACAAACAGGCGCCAGGACATCCCGGATACCT 2069
291 yAlaLeuThrAlaGluValGlyAspMetThrGluAspSerAlaGlnAla 308
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2070 AAGACCACTTCTAGAGATCAAAAGCATTTCCCTTTATATATGCTCTGA 2119
308 euIleArgGluValLeuIleLysGluLeuAla.....GluGlyLeu 322
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2120 TATTAAGGAAGCTCTTTTCATATATATGATGCTGTGGAAAGTTTATG 2169
323 LeuGlnIleGlnLysSerArgCysLeuArgAsnLeuMetLysThrPro 339
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2170 GTTTACTTTGGAAAGAACCAAGTTTGCAGAGATACAAAGAAATCTCTCT 2219
339 uPheValValIleThrCysAlaIleGlnMetGlyLysSerGluPheHis 356
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2220 CTTTGTGGCGGAGTCTGTCTCATTTGTTCAATTCCTTTGACCCAT 2269
356 eHisIleThrGlnThrLeuPheHisThrPheTyrAspLeuIleGln 372
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2270 CCTTGATGATATGGGCTGTTTCAAGTCTCTATATGAGACGCTTTCCTTA 2319
373 LysAsnLysHisLysHisLysGlyValAlaIleSerAspPheIleArg.. 388
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2320 AGCAACAA.....GCGACAGCTGAATTTCTCAAAAC 2351
389 ..SerLeuAspHisCysGlyAspLeuAlaLeuGlnGlyValPheSerHisL 405
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2352 AACTGTCTCTCTGTGAGCTGCTTGAAGGCTTTTTCATGTT 2401
405 yspPheAspPheGluLeuGlnAspVal.....SerSerValAsnGluAsp 419
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2402 GCTTGAAGTTTATGATGATGATCTCGCAGAACAGGCTTATGATAGAT 2451
420 ValLeuLeuThrThrGlyLeuLeuCysLysTyrThrAlaGlnArgPheLys 436
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2452 GAAGATCTAACCATGCTGCTTGAATGAGCAAAATTTACAGCCCAAGACTAAG 2501
436 sProLysTyrLysPhePheHisLysSerPheGlnGluTyrThrAlaGly 453
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2502 ACCATTCTACCGGTTTTAAGTCTGCTTCCAGAGATTTCTTGGGGGA 2551
453 rArgLeuSerSerLeuLeuThrSerHisGluProGluGluValThrLys 469
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2552 TTAGGCTGATGTAACCTCGCATTCAGATAGCCAGGACATCAAGATTGG 2601
470 GlyAsnGlyTyrLeuGlnLysMetValSerIleSerAspIleThrSerTh 486
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2602 GCACGTATACATTTGAACAAATCAACCTCACCAGATGACTGTAAAGGC 2651
486 rTyrSerSerLeuLeuArgTyrThrCysGlySerSerValGluAlaThr 503
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2652 CTACAAACATTTTGAATATGTC.....TCCAGCTCTCCCTTAACAA 2695
503 rGAla.....ValMetLysHisLeuAlaIleValTyrGlnHisGly 516
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2696 AAGCAGGGCCCAAAATGTGTCTCATTTTGTCTCATTTAGTGATTAAC... 2742
517 CysLeuLeuGlyLeuSerIleAlaLysArgProLeuTyrPargGlnGly 533
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2743 .....AAAGAGTC 2750
533 rLeuGlnSerValLysAsn..... 539
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2751 ATTGAGAGATATATCTGAATATGATGACTTTAAAGACAGACAGCAAA 2800
540 ..ThrThrGluGlnGluIleLeuLysAlaIle.....AsnIle 551
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

2801 TTTCACGACGATGCACTTACTTAGGGGATTGCGCAATTTGTCACAA 2850
552 AsnSerPheValGluCysGlyIleHisLeu..... 561
2851 GCTTACTTTCATGTTTCAGAACATTTACTGTCTTGCCCTGAAAC 2900
562 ...TGTGInGluSerThrSerLysSerAlaLeuSerGlnIuPheGlu 577
2901 TGCTTTCAGAACGACACT...GTTCGTGCGTGTTCATTTGTTTGGC 2947
577 IaphePheGlnGlyLysSerLeuThrIleAsnSerGlyAsnIleProAsp 593
2948 AATTCCCTTCAGGAGAACACTACTTGGGTGGCTTAACCTTA...CAG 2994
594 TyrLeuPheAsp..... 597
2995 TACTTTTTCAGCCACCAGAAAGCTGTCTATTTGTAGAGACATCCACT 3044
598 PhePheGlnHisLeuP 603
3045 CCCAATACGAGAAATTAAGACATCCACCAGACACATTTTCACTTCTG 3094
603 roAsnGly.....AlaSerAlaLeuAspPheIleLysLeuAspPheTyr 617
3095 AAACATGTTTTCAGAACATTCACAGTCCCACTATAGATCAGGACTATGCT 3144
618 GlyGly.....AlaMetAlaSerTyrGlu..... 625
3145 TGTGCTTTGAACTTATGATGATGGAGCGAAATTTAGCTGAAAGA 3194
626 LysAlaAlaGluAspT 631
3195 GGATATGTAAGAGCTATATGATATGACGCGAGGCGATCCACCAAGC 3244
631 hrGlyGlyIleHisMetGluGluAlaProGlnThrTyr...IleProSer 646
3245 TTAGTACTGGCTATTTGGAACCTTCTCCAAAGCGTACAGATTC... 3291
647 ArgAlaValSerLeuPhePheAsnTyrLysGlnIuPheArgThrLeuG 663
3292TGTCTAGA 3299
663 ValAlThrLeuArgAspPheSerLysLeuAsnLysGlnAspIleThrTyrL 680
3300 AGTCGATGTAATGATATGATGTTGTAGGCGAGATATGCTTAGATTC 3349
680 euLysLysIlePheSerSerAlaThrSerLeuArgGlnIleLysArg 696
3350 TAAATGACAGTTTCTCTGAGTTCACAGCGCATCGAATTCATTTAAACAC 3399
697 CysAlaGlyValAlaGlySerLeuSerValLeuSerThrCysLys.. 712
3400 AGCAGAGGCTTTATAGAAAGCATCCGCCAGCTCTGAGCTGTCTAGGC 3449
713 .AsnIleTyrSerLeuMetValGluAlaSerProLeuThrIleGluAspG 729
3450 CTCTGTACCAAGTCTCCATTAAGCAATTGGAATCAGCGCAGCCGAAAC 3499
729 IuArgHisIleThrSerValThrAsnLeuLysThrLeuSerIleHis... 744
3500 AGAAGACTGCTTCACCCCTGCTCTCTGGAATCTCTGGAAGTCTCAGG 3549
745AspLeuGlnAsnGlnArgLeuProGlyLysLeuThrAspSerLe 759
3550 ACAATCCAGTCACAGACCAAAATCTTCTTAAT...CTGCAATAGTT 3593
759 uGlyAsnLeuLysAsnLeuThrLysLeuIleMetAspAsnIleLysMet. 775
3594 CCTGTGCTGAAGAAGACTGTCTGTGATGTGAGGAGCAATATTAATGTTT 3643
775 775
3644 TTTCAGTCAATCTCGAAGAATTTCCAAACTTCCACCATATGAGAAATTA 3693

776AsnGluGluAspAlaIleLysLeuAlaGlyLe 787
3694 TTGATCCAAATTTGACGTAGATATGCTTCCAAACTAGTAATAATAT 3743
787 uLysAsnLeuLysLysMetCysLeuPheHisLeu.....ThrHisLeu 802
3744 TCAAAATCTCCAAACCTCATGTTTTCATCTGATCTGATTAATCTCTTT 3793
802 erAspIleGlyGluIleLysAspTyrIleValLysSerLeuSerSerGlu 818
3794 CGGATTTGGGTCTCTGACTACTGCTGTTTCC..... 3828
819 ProGly...AspLeuGluGluIleGlnLeuValSerCysCysLeuSerAl 834
3829 ...TGTAAGAAACTCAGAGAAATTAAGTTTCGGATTCATTTT... 3870
834 aAsnAlaValLysIleLeuAlaGlnAsnLeuHisAsnLeuValLysLeu 851
3871 .CAAGCCGTCCCATTTGTGCC...AGTTGCCAAATTTATTTCTCTGA 3916
851 erIleLeuAspLeu...SerGluAsnTyrLeuGlnLysAspGlyAsnGlu 866
3917 AGATATTAATCTTGAAGGCCAGCAATTTCTCTGATGAGAAACATTCAGA 3966
867 AlaLeuHisGluLeuIleAspArgMetAsnValLeuGlnIleuThrAl 883
3967 AAATTTGCTTACATTTTAGGTTCTCTTAGTAACCTGGAAGA..... 4008
883 aLeuMetLeuProTyrGlyCysAspValGlnGlySerLeuSerSerLeu 900
4009 .TTGATCTTCTTACTGCGGATGGAATTTATGAGTGGCCAAACTGATCA 4057
900 euLysHisLeuGluGluValProGlnLeuValLysLeuGlyLeuLysAsn 916
4058 TCCAGAGGTGCACAGCTTCATGTCTCCAGTCCCTCATTTTCAAG 4107
917 TyrArgLeuThrAspThrGluIleArgIleLeuGly..... 928
4108 ...ACTTGAATGATGACAGCGTGTGGAATTTGTGAGCTAGTGTTC 4154
929Alap 930
4155 GCTTGCAATGCAAGCAGTGTATAGCCAGCTTCTGCTGCACATGTCT 4204
930 hePheGlyLysAsnProLeuLysAsnPheGln.....Gln 941
4205 ATGTAAACATTTGCCCTTAGAATAATTTTCAACCCGCTCTCATTTTCA 4254
942 LeuAsnLeuAlaGlyAsnArgValSerSerAspGlyTyrLeuAlaPhe 958
4255 CTATCATACGTGTTCTCTGATGTCCTGTGTGATTTAGGCGGATTCGTG 4304
958 tGlyValPheLeuAsnLeuLysGlnLeuValPhePheAspPheSerThr 974
4305 GTCAGATTTGGAAGTACAAAAGGTCTCCATTTGTGTGATATACAAAGCC 4354
975LysGluPheLeuProAspProAlaLeu 983
4355 TCAAAATCTGCGTTCTTCCACACTGTGTTTGA 4386
seq_name: /STD2/gcgdata/geneseq/geneseqn/NA1999.DAT:AAK56273
seq_documentation_block:
ID AAK56273 standard; cDNA to mRNA; 5366 BP.
XX
AC AAK56273:
XX
DT 20-JUL-1999 (first entry)
XX
DE Human apoptosis inhibiting protein encoding cDNA #2.
XX
KW Human; apoptosis inhibitory protein; apoptotic disease; diagnosis;

KW spinal muscular atrophy; ds.
XX Homo sapiens.
OS JP1116599-A.
XX 27-APR-1999.
PD 14-OCT-1997; 97JP-0280831.
XX 14-OCT-1997; 97JP-0280831.
PR 14-OCT-1997; 97JP-0280831.
XX (KAGA-) KAGAKU GIUTSU SHINKO JIGYODAN.
PA WPI: 1999-323531/27.
DR P-PSDB; AAY09540.
XX
XX
XX New apoptosis inhibitory protein - useful for determining mechanism
PT of various apoptotic diseases e.g. human spinal muscular atrophy
XX
XX
PS Claim 8; Page 13-15; 16pp; Japanese.
CC The present sequence encodes a human apoptosis inhibitory protein.
CC The apoptosis inhibitory protein is useful for the elucidation of
CC the mechanism of various apoptosis diseases such as human spinal
XX muscular atrophy and the diagnosis, the prevention and the treatment
XX of such diseases.
SQ Sequence 5366 BP; 1483 A; 1173 C; 1222 G; 1488 T; 0 other:

alignment_scores:
Quality: 593.50 Length: 1094
Ratio: 1.023 Gaps: 42
Percent Similarity: 53.016 Percent Identity: 23.675

alignment_block:
US-09-697-089-2 x AAY56273 ..

Align seg 1/1 to: AAY56273 from: 1 to: 5366

69 pheLeuLysSerLeuLysGluTrpAsnTyrProLeuPheGlnAspLeuAs 85
1321 TTTCTCCAAATATGAAATCC...TCTGGGAAATGATCCACCACTTCA 1367
-85 nclYgln.....SerLeuPheHisGlnThrSerGluGlyAspL 98
1368 GAGCCGTGGTGAACCTTGTGAATTACTGGAACCAAGTGAAGCAATC 1417
98 euAspAsp..... 100
1418 TTGAAATTCATAGCAGTTGGTCTATAGTCCAGAAATGGCAGCGGT 1467
101LeuAlaGlnAspLeuLysAs 107
1468 GAGCCAGTGGTTTCAAGAGCAAGAAATGCAATGATGAGACGTGAGAG 1517
107 pleuTYrHisThrProSerPheLeuAsnPhe..... 117
1518 AGCTTATACAGCGCATTTCCGCCACATGTCCTTGGTATATCTCTT 1567
118TyrProLeuGlyGluAspLeuAspLeuPhe 128
1568 CCGATGTGGCCAGCAGACCTTCTGCGCTGTGATCTGTCTATT..... 1611
129 AsnLeuLysSerThrPheThrGluProValLeuTyrPArgLysAspLnh 145
1612 ...GCTTCAAAACACATCAAGAAACCTGTG..... 1638
145 sHisHisArgValGluGlnLeuThrLeuAsnGlyLeuLeuGlnAlaLeuG 162
1639CAAGAACCTCTGCTGCTGCTGAGTCTTTGGCACTTGA 1678

162 InSerProCysIleIleGlnGlyLysSerGlyLysGlyLysSerThrLeu 178
1679 ACTGTCTCATGTGTGTGAGGTTGAAGCTGGAGTGAAGTGAAGAGCGTCTC 1728
179 LeuGlnArgIleAlaMetLeuTrpGlySerGlyLysCysLysAlaLeuTh 195
1729 CTGAAGAAATACCTTTTGTGTGGCATCTGGATGCTGCCCTGTAA 1778
195 rLysPheLysPheValPhePheLeuArgLeuSer.....ArgAlaGln 210
1779 CAGGTTCCAGCTGGTTTCTTACCTCCCTTAACTCCACAGCAGCAGAG 1828
210 LysLeuPheGluThrLeuCysAspGlnLeuLeuAspIleProGlyThr 226
1829 AGGGGCTGGCCAGTATCATCTGTGACCACTCTTAAGAAAGAAAGATCT 1878
227 IleArgLysGlnThrPheMetAlaMetLeuLeuLysLeuArgGlnArg 243
1879 GTTACTGAATGTGCATGAGGAACATTATCCACGACTTAAAGAAATCAG 1928
243 IleuPheLeuLeuAspGlyTyrAsnGlnPhe.....LysProGlnAsn 258
1929 CTATATCTCTTTAGATGACTACAAAGAAATATGTCATCCCTCA..... 1974
258 LysProGluIleGlnAlaLeuIleLysGluAsnHisArgPheLysAsnMet 274
1975GTCAATAGGAACCTGATTCATAAAACCACTTATCCCGGACCTGC 2019
275 ValIleValThrThrThrThrGluCysLeuArgHisIleArgGlnPhe 291
2020 CTATGATGCTGTGCTCCGTACAAACAGGCGCAGGACATCCGCGATACCT 2069
291 yAlaLeuThrAlaGluValGlyAspMetThrGluAspSerAlaGlnAla 308
2070 AGAGACCAATCTAGAGATCAAAACCAATTCCTTTTAAATAGTGTCTGA 2119
308 euIleArgGluValLeuIleLysGluLeuAla.....GluGlyLeuLeu 322
2120 TATTACGGAAGCTCTTTTCAATATATGATGCTGTGCGAAAGTTTATG 2169
323 LeuGlnIleGlnLysSerArgCysLeuArgAsnLeuMetLysThrPro 339
2170 GTTACTTGTGGAAGAACCAAGTTTGCAGAAATGATCAGAAACCTCTCT 2219
339 urPheValValIleThrCysAlaIleGlnMetGlyLysSerGluPheHis 356
2220 CTTTGTGGCGCGCATCTGTCTCATTTGTTGTTTGGACCAT 2269
356 eHisThrGlnThrThrLeuPheHisThrPheTyrAspLeuLeuIleGln 372
2270 CCTTTATGATGATGGCTGTTTCAAGTCCATATGAGACGCCCTTCTTA 2319
373 LysAsnLysHisLysHisLysGlyValAlaAlaSerAspPheIleArg.. 388
2320 AGGAACAAA.....GCGACAGCTGAATTCATCAAGC 2351
389 .SerLeuAspHisCysGlyAspLeuAlaLeuGluGlyValPheSerHisL 405
2352 AACTGTGTCCTCTGTGTGAGCTGGCTTGAAGGGTTTTTTCATGCTT 2401
405 LysPheAspPheGluLeuGlnAspVal.....SerSerValAsnGluAsp 419
2402 GCTTTAGTTTAATGATGATGATCTGCGAAGCAGGAGGTGATGAAGAT 2451
420 ValLeuLeuThrThrGlyLeuLeuCysLysTyrThrAlaGlnArgPhe 436
2452 GAAGATCTAACCATGTGCTTGAATGAGCAAAATTAACAGCCAGAGACTAAG 2501
436 sProLysTyrLysPhePheHisLysSerPheGlnGluTyrThrAlaGlyA 453
2502 ACCATTCTACCGGTTTTTAAAGTCCCTGCCATCAAGAAATTTCTGGCGGA 2551
453 rGArgLeuSerSerLeuLeuThrSerHisGluProGluGluValThrLys 469

```

2552 TGAGCGTGATTGAGACTCTCGATTGCAGATAGCAGCAACATCAAGATTG 2601
470 GlysAsnGlyThrLeuGlnIleuMetValSerIleSerAspIleThrSer 486
2602 GGACTGATCATTTGAAACAATCACTACCCATGATGACTGATGAAGCC 2651
486 rTySerSerLeuLeuArgThrCysGlySerSerValGluIleThr 503
2652 CTACACAATTTTGAACATATGTC.....TCCAGCCTCCCTTCAACAA 2695
503 rGluA.....ValMetLysHisLeuAlaIleValTyrGlnHisGly 516
2696 AAGCAGGCGCCAAATTTGTCTCATTTGCTCCATTAGTGGAATAC... 2742
517 CysLeuLeuGlyLeuSerIleAlaIleArgProLeuThrArgGlnIle 533
2743 .....AAAGAGTC 2750
533 rLeuGlnSerValLysAsn..... 539
2751 ATTGAGAAATATATCTGAAATGATGACTACTTAAGCACCAGCAAA 2800
540 ..ThrThrGlnGlnIleLeuLeuAlaIle.....AsnIle 551
2801 TTTCACCTGCAGATCGAGTTACTTAGGGGATTGTGCAAATTTGTCCAA 2850
552 AsnSerPheValGluCysGlyIleHisLeu..... 561
2851 GCTTACTTTCAATGTTTCAGACATTTACTGTTCTTGCCCTGAAAC 2900
562 ....TyrGlnGluSerThrSerLysSerAlaLeuSerGlnIlePheGlu 577
2901 TGCTTATCAAGCAACACT...GTTGCTGCGTGTCTCCATTGTTTTC 2947
577 LaspheGlnGlyLysSerLeuTyrIleAsnSerGlyAsnIleProasp 593
2948 AATTCCTTCAAGGAGAACCTGACTTGGGTGGCTTAACCTTA...CAG 2994
594 TyrLeuPheasp..... 597
2995 TACTTTTCGACCCAGAAAGCTTGCATTGTGAGGACATCCACTT 3044
598 .....PhePheGlnHisLeuP 603
3045 CCCAATACGAGAAATAGACATCACCCAGACACATTTTCAGTTCTCG 3094
603 roAsnGly.....AlaSerAlaLeuAspPheIleLysLeuAspPheTyr 617
3095 AATCATGTTTGCACAATTCACAGGTGCCACACTATATGATCAGGACTATGCT 3144
618 GlyGly.....AlaMetAlaSerTrpGlu..... 625
3145 TCTGCCCTTGAACCTATGATGATGGAGGAAATTTAGCTGAAAAAGA 3194
626 .....LysAlaAlaLeuAsp 631
3195 GGAAATATGTAAGAGCTATATGATATGACCGCAGGGGACATCCACAGACC 3244
631 hrGlyGlyIleHisMetGlnGluIleProGlnIleThrTyr...IleProSer 646
3245 TTACGTACTGGCTATGGAACCTTCTCCAAAGCAGTACAGATTC... 3291
647 ArgAlaValSerLeuPheAsnTrpLysGlnIlePheArgThrLeu 663
3292 .....TGTCTAGA 3299
663 uValThrLeuArgAspPheSerLysLeuAsnLysGlnAspIleThrTyrL 680
3300 AGTCGATGTAATGATATGATGTTGAGCCAGCATATGCTTGAGATTC 3349
680 euGlyLysIlePheSerSerAlaThrSerLeuArgLeuGlnIleLysArg 696
II .....
3350 TAAATGACAGTTTCTCAGCTTCACAGCGCATGCAACTCCATTAAACAC 3399
697 CysAlaGlyValAlaGlySerLeuSerValLeuSerThrCysLys.. 712
3400 AGCAGAGCGCTTATAGAAAGCATCCGCCAGCTCTTAGCTGTCTAAGC 3449
713 .AsnIleTyrSerLeuMetValGluAlaSerProLeuThrIleGluAsp 729
3450 CTCTGTCCACCAAGTCTCCATACCAAGTTCGAACCTCAGCCAGCCGAC 3499
729 LuArgHisIleThrSerValThrAsnLeuLysThrIleSerIleHis... 744
3500 AGGAATGCTCTTCACCTGCTGCTGGAATCTTGAAGTCTCAGCG 3549
745 .....AspLeuGlnAsnGlnArgLeuProGlyGlyLeuThrAspSerLe 759
3550 ACATTCACGTCCACAAAGCAATCTTCTCTAAT.....CTGAAATAGTT 3593
759 uGlyAsnLeuLysAsnLeuThrLysLeuIleMetAspAsnIleLysMet 775
3594 CCTGTGCTGCAAGAAAGACTGCTGTGATCTGAGGCGCAATTAATGTTT 3643
775 ..... 775
3644 TTTCAGTCATTCCTGAGCAATTTCCAAACTTCACACCATATGAGCAATTA 3693
776 .....AsnGluGluAspAlaIleLysLeuAlaGlyLe 787
3694 TTGATCCAAATTTACAGTAGATATGCTTCACAACTAGTAATATATAT 3743
787 uLysAsnLeuLysLysMetCysLeuPheHisLeu.....ThrHisLeu 802
3744 TCAAAATTTCTCCAAACCTTCATGTTTCCAACTGAAAGTGAATCTTCTT 3793
802 erAspIleGlyGlyLeuMetAspTyrIleValLysSerLeuSerGlu 818
3794 CGGATTTGGGTCTCTCATGACTATGCTTGTTC... 3828
819 ProCys...AspLeuGluGluIleGlnLeuValSerCysCysLeuSerAl 834
3829 ...TGTAAGAACTCCAGAAATTAAGTTTCGGATCATTTT... 3870
834 AsnAlaValLysIleLeuAlaGlnAsnLeuHisAsnLeuValLysLeu 851
3871 .CAAGCCCTCCCATTTGTGCC...AGTTGCCAAATTTATTTCTCTCGA 3916
851 erIleLeuAspLeu...SerGlnAsnTyrLeuGlnLysAspGlyAsnGlu 866
3917 AGATATTAATCTTGAAGCCAGCAATTTCTGATGAGAAACATCAACA 3966
867 AlaLeuHisGluLeuIleAspArgMetAsnValLeuGlnIleuThrAl 883
3967 AAATTTGCGCTACATTTAGGTTCTTATGTAACCTGGAAGAA... 4008
883 aleuMetLeuProTrpGlyCysAspValGlnGlySerLeuSerSerLeu 900
4009 .TTGATCCTTCTACTGCGGATGGAATTTATGAGTGGCCAAACTGACGA 4057
900 euLysHisLeuGluGluValProGlnLeuValLysLeuGlyLeuLysAsn 916
4058 TCCAGCAGTGCACAGCTTATGTTCTCCGAGTCCCTCTCATTTTTCAG 4107
917 TrpArgLeuThrAspThrGlnIleArgIleLeuGly..... 928
4108 ...ACTTGATGATGACAGCGGTGTGGAATTTGCTGAGTACTGTTTCA 4154
929 .....Alap 930
4155 GCTTGCAATGGAAGCCAGTGTATAGCCAAAGCTTTGTGTCGACATGCT 4204
930 hePheGlyLysAsnProLeuLysAsnPheGln.....Gln 941
4205 ATGTAACATTTGCCCTTAGAAATTTTCAACCCGCTTCTCATTTTCA 4254

```

```

942 LeuAsnLeuAlaGlyAsnArgValSerAspGlyTrpLeuAlaPheX 958
      |||
      |||
      |||
4255 CATCATACTGTTCTTCTAGTCTCTGTGCAATTAGGGCATCTCG 4304
      |||
      |||
      |||
958 tGlyValPheGluAsnLeuValPheAspPheSerThr. 974
      |||
      |||
      |||
4305 GTCAGATTGGAGTACAAAGAGTCTCCATTTGTGATATACAAAGCC 4354
      |||
      |||
      |||
975 .... LysGluPheLeuProAspProAlaLeu 983
      |||
      |||
      |||
4355 TCAGATTGCGTTCTTCCACCTGCTGTTTA 4386
      |||
      |||
      |||
seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT:AAT30092
seq_documentation_block:
ID AAT30092 standard; cDNA: 5502 BP.
XX
XX AAT30092:
XX
XX 30-DEC-1996 (first entry)
XX
XX Neuronal apoptosis inhibiting protein coding sequence.
XX
XX Neuronal apoptosis inhibiting protein; human; NAIP; chromosome 5q13; YAC;
XX yeast artificial chromosome; spinal muscular atrophy; mammalian cell;
XX autosomal recessive; neurodegenerative disorder; alpha motor neuron; SMA;
XX spinal cord; proximal voluntary muscle; therapy; apoptotic mechanism; ss.
XX Homo sapiens.
XX
XX OS
XX
XX Key Location/Qualifiers
XX FH 396..4094
XX CDS /*tag= a
XX FT /product= neuronal apoptosis inhibiting protein
XX
XX WO9612016-A1.
XX
XX 25-APR-1996.
XX
XX 17-OCT-1995; 95MO-CA00581.
XX
XX 19-DEC-1994; 94CA-2138425.
XX
XX 18-OCT-1994; 94GB-0012019.
XX
XX (SHKJ ) RES DEV CORP JAPAN.
XX (OTOT-) UNIV OTTAWA.
XX
XX Ikeda J, Korneluk RG, Mackenzie AE, Mahadevan MS;
XX McLean M, Roy N;
XX
XX WPI: 1996-222003/22.
XX P-PSDB; AAR89217.
XX
XX Neuronal apoptosis inhibitor protein gene - used to develop prods.
XX for use in the diagnosis and therapy of spinal muscular atrophy
XX
XX Claim 4; Page 64-67; 113pp; English.
XX
XX This sequence represents the cDNA sequence for the human neuronal
XX apoptosis inhibitor protein (NAIP). This sequence was found on a region
XX of the human chromosome 5q13. This sequence was isolated from a yeast
XX artificial chromosome (YAC) contig containing the D5S435-D5S112 interval
XX of the chromosome 5q13. Mutations in this gene, are causative of spinal
XX muscular atrophy (SMA) types I, II, and III. SMAs are a group of
XX autosomal recessive, neurodegenerative disorders. SMAs are classified
XX into the three types based upon the age of onset (with type I being the
XX severest form with the earliest age of onset). All three types are
XX characterised by the degeneration of the alpha motor neurons of the
XX spinal cord manifesting as weakness and wasting of the proximal voluntary
XX muscles. The most common mutations of this sequence are thought to be
XX deletions of exons 5 and 6, and reductions in the copy number of the
XX gene. This gene, (and primers and probes based on this gene) can be used

```

```

CC for the diagnosis of SMA, and for directing the formulation of
CC conventional and genetic therapies for SMA. Identification of genes
CC showing homology with the NAIP locus, and proteins that interact with
CC NAIP can be used in the elucidation of apoptotic mechanisms in mammalian
CC cells.
XX
XX Sequence 5502 BP; 1558 A; 1229 C; 1204 G; 1511 T; 0 other;
XX
XX
XX alignment_scores:
XX      Quality: 561.50      Length: 937
XX      Ratio: 1.123      Gaps: 33
XX Percent Similarity: 53.362      Percent Identity: 24.120
XX
XX alignment_block:
XX US-09-697-089-2 x AAT30092 ..
XX
XX Align seg 1/1 to: AAT30092 from: 1 to: 5502
XX
XX 69 PheLeuLysSerLeuLysGluTrpAsnTrpProLeuPheGlnAspLeuAs 85
XX |||
XX |||
XX 1425 TTCTCCAAAATATGAGATCC...TCTGGGAGAGTCCAGACCTTCA 1471
XX
XX 85 nGlyGln.....SerLeuPheHisGlnThrSerGluGlyAspL 98
XX |||
XX |||
XX 1472 GAGCCGTGTTGACTTGTGATTTACTGMAACCAACCAAGTGAACATC 1521
XX |||
XX |||
XX 98 euAspAsp.....
XX |||
XX |||
XX 1522 TTGAAGATTCAATAGCAGTGGCTCTATAGTCCAGAAATGGCACAGG 1571
XX
XX 101 .....LeuAlaGlnAspLeuLysAs 107
XX |||
XX |||
XX 1572 CAAGCCAGTGGTTTCAAGAGGCAAGATCTGAATGAGCAGCTGAGAGC 1621
XX
XX 107 PLeuTrpHisThrProSerPheLeuAsnPhe..... 117
XX |||
XX |||
XX 1622 AGCTTATACACAGCGCAGTTTCCGCCACATGCTTGGTGAATATCTT 1671
XX
XX 118 .....TyrProLeuGlyGluAspIleAspIlePhe 128
XX |||
XX |||
XX 1672 CCGATCTGGCCAGCAGCACCTTGTGGCTGTGATCTGTATTT..... 1715
XX
XX 129 AsnLeuLysSerThrPheThrGluProValLeuTrpArgLysAspGlnH 145
XX |||
XX |||
XX 1716 ..GCTTCAAAACACATCAGCAAACTGTG..... 1742
XX
XX 145 SHisHisArgValGluGlnLeuThrLeuAsnGlyLeuLeuGlnAlaLeu 162
XX |||
XX |||
XX 1743 .....CAAGAACCTCTGGTCTGCTCCAGAGCTTTTGGCAACTGA 1782
XX
XX 162 InSerProCysIleIleGluGlyLysSerGlyLysGlyLysSerThrLeu 178
XX |||
XX |||
XX 1783 ACTGTGTCATGTGTGTGGAGGCTGGAAGCTGGAAGCAAGCAGCTCTC 1832
XX
XX 179 LeuGlnArgGluIleAlaMetLeuTrpGlySerGlyLysCysLysAlaLeu 195
XX |||
XX |||
XX 1833 CTGAAGAAATAGCTTTCTGTGGGCACTGTGATCTCCCTGTAA 1882
XX
XX 195 rLysPheLysPheValPhePheLeuArgLeuSer.....ArgAlaGln 210
XX |||
XX |||
XX 1883 CAGGTTCAGAGCTGTTTCTACTCTCCCTTAGTCCACAGACAGACAGC 1932
XX
XX 210 lGlyLysPheGluThrLeuCysAspGlnLeuLeuAspIleProGlyThr 226
XX |||
XX |||
XX 1933 AGGGCTGCGCATCATCTGTGTGACAGCTCTGAGAAAGAGAGACT 1982
XX
XX 227 lLeaGlyLysGlnThrPheMetAlaMetLeuLysLeuArgGlnArgVa 243
XX |||
XX |||
XX 1993 GTTACTGAATGTGCATGAGAACATTTATCCAGCAGTTAAAGATCAAG 2032
XX
XX 243 lLeuPheLeuLeuAspGlyTyrAsnGluPhe.....LysProGlnAsn 258
XX |||
XX |||
XX |||

```



```

775 ..... 775
3748 TTTCAGTCAATCTCTGAGAAATTTCCAACTTCCACATATGGAATA 3797
776 ..... 787
3798 TTGATCCAAATTTACAGTATGATCTCTCCAACTGATGCCAGTTT 3847
787 uylsasnlleuylslysmetCysleuphehlsleuthrlsleu..... 801
3848 GCCAAATTTTATTTCTCTGAAGATATTAATCTTAAGGCCACCAATTC 3897
802 ..... Seraspilleglygluylmetaspyrillevalysserleuser 816
3898 CTGATGAGAGAACATCGAAAAATTTGCCATATTAGTTCTCTTACT 3947
817 SerGluproCysaspileuylgluylleuvalserCysCysleuse 833
3948 ..... AACCTGGAGAAATGATCTCTCTACTGGGATGGAAT 3985
833 ralasnaAlaValylsIleleuAlaGlnAsnleuHisAsnleuValLysL 850
3986 TTATCGAGTGGCCAACTGATCATCGACGATGTCAGAGCTTCATTTGC 4035
850 euserlleu 853
4036 TCCGAGTCTCTC 4046

```

seq_name: /SIS2/gcdata/geneseq/geneseq/NA197.DAT:AA71263

seq_documentation_block:

ID AA71263 standard; cDNA; 5502 BP.

AC AA71263:

DT 06-OCT-1997 (first entry)

DE Neuronal apoptosis inhibitor protein (NAIP) cDNA (no exon 14a).

KW Neuronal apoptosis inhibitor protein; NAIP; diagnosis;

KW gene therapy; cancer; AIDS; amyotrophic lateral sclerosis;

KW spinal muscular atrophy; ss.

OS Homo sapiens.

PN W09726331-A2.

PD 24-JUL-1997.

PF 17-JAN-1997; 97WO-IB00142.

PR 19-JAN-1996; 96GB-0001108.

XX (UYOT-) UNIV OTTAWA.

PI Korneluk RG, Mackenzie AE, Robertson G, Roy N, Tamai K;

DR WPI: 1997-385335/35.

PT New neuronal inhibitor of apoptosis - useful for diagnosing and

PS treating, e.g. cancer, AIDS or amyotrophic lateral sclerosis

XX Example 4; Fig 5A-L; 102pp; English.

CC A new neuronal apoptosis inhibitor protein (NAIP) cDNA clone

CC (AA71263) was isolated from a human foetal spinal cord cDNA library

CC by probing with the genomic insert in cosmid 25086, containing a

CC CAPT locus. An additional coding sequence (AA71264), including a

CC exon 14a (see also AA71265 and AA71266), was subsequently obd.

CC The NAIP DNA sequence including exon 14a appears to be a

CC predominant gene isoform which is not deleted or mutated in spinal

CC muscular atrophy (SMA) patients. The NAIP gene was mapped to

CC 5q13.1. NAIP (see also AAW20032 and AAW20033) is a negative regulator

CC of apoptosis, partic. neuronal apoptosis and, when deficient or
 CC absent, contributes to neurodegenerative phenotypes such as SMA and
 CC amyotrophic lateral sclerosis.

XX SQ Sequence 5502 BP; 1560 A; 1226 C; 1206 G; 1510 T; 0 other;

alignment_scores:

Quality:	546.50	Length:	950
Ratio:	1.078	Gaps:	38
Percent Similarity:	53.368	Percent Identity:	24.000

alignment_block:

US-09-697-089-2 x AA71263 ..

Align seg 1/1 to: AA71263 from: 1 to: 5502

69 PheLeuLysSerLeuLysGlnTrpAsnTyrProLeuPheGlnAspLeuAs 85

1425 TTCTCTCAAAATATGAAATGTC...TCTCGGAAGTACTCCAGACCTTCA 1471

85 nglYgln.....SerLeuphehlsGlnThrSerGlulYAspI 98

1472 GAGCCCTGGTGAACCTTGTGAATTAAGTGAACCAAGTGAAGCAATC 1521

98 euAspAsp..... 100

1522 TTGAAGATTCATATGACGATTGCTATATAGTCCAGAAATGCGACAGGT 1571

101LeuAlaGlnAspLeuLysAs 107

1572 GAAGCCAGTGGTTTCAAGAGGCAAGAAATATGAAATGACAGCTGAGAC 1621

107 pleuTYrHisThrProSerPheLeuAspPhe..... 117

1622 AGCTTATACAGCGCGATTTCCGCCACATGCTTTCCTGTGATNCTCTT 1671

118TyrProLeuGlyGlnAspIleAspIlePhe 128

1672 CCGATTTGGCCAGCGACCACTTGGCTGTGATGTGCTATTT..... 1715

129 AsnLeuLysSerThrPheThrGluProValLeuTrpArgLysAspGlnH 145

1716 ...GCTTCAAAACACATCAGCAAACTGTG..... 1742

145 SHSHSHArGValGluGlnLeuThrLeuAsnGlyLeuLeuGlnAlaLeuG 162

1743CAAGAACCTCTGCTGCTGCTGAGGCTTTGGCAACTTGA 1782

162 InSerProCysIleIleGluGlyGlnSerGlyLysGlyLysSerThrLeu 178

1783 ACTCTGTCAATGTGTGAGAGGTGAAGCTGGAAGTGAAGAGAGGCTCTC 1832

179 LeuGlnArgIleAlaMetLeuTrpGlySerGlyLysCysLysAlaLeuTh 195

1833 CTGAAGAAATAGCTTTTCTGTGGGCATCTGATGTGCTCCCTGTAA 1882

195 rlySpheLysPheValPhePheLeuArgLeuser.....ArgAlaGlnG 210

1883 CAGGTTCCAGCTGCTTTTCTACTCTCCCTTAGTTGCACAGCAGAGAG 1932

210 LylGlyLeupheGlnThrLeuCysAspGlnLeuLeuAspIleProGlyThr 226

1933 AGGGCTGGCCAGTATCATCTGTGACCACTCTCTAGAGAAAGAGATCT 1982

227 IleArgLysGlnThrPheMetAlaMetLeuLeuLysLeuArgGlnArgVa 243

1983 GTTACTGAATGATGATGAGAAATATTATTCAGCAAGTAAAGATACGCT 2032

243 lleuPheLeuLeuAspGlyTyrAsnGluPhe.....LysProGlnAsnC 258

2033 CTATATCTTTAGATGACTACAAAGAAATATGTTCAATCTCTCA... 2078

```

258 yspGluIleGluAlaLeuIleLysGluAsnHisArgPheLysAsnMet 274
    ||| |||||:|||||
2079 ....GTCATGGAAGAACTGATCAAAAAACCATTTATCCCGACCTGC 2123
275 ValIleValThrThrThrThrGluCysLeuArgHisIleArgGlnPheG1 291
    :|||:|||||:|||||:|||||
2124 CTATTGATGTGTCGCTACAAACAGGCGCAGGACATCCCGCATTTACT 2173
291 yAlaLeuThrIaGluValGlyAspMetThrGluAspSerIleAlaIle 308
    |||:|||||:|||||:|||||
2174 AGAGACATCTTAGATCAAGATCAAGATTTCCCTTTATATATCTGTCTA 2223
308 euIleArgGluValLeuIleLysGluLeuAla.....GluGlyLeuLeu 322
    :|||:|||||:|||||:|||||
2224 TATTACGGAAGCTCTTTACATATATATGACTCTGCGAAGTTTATG 2273
323 LeuGlnIleGlnLysSerArgCysLeuArgAsnLeuMetLysThrProle 339
    :|||:|||||:|||||:|||||
2274 GTTACTTTGGAAAGAACCAAGTTTGCGAAGATACGAAAGACTCTCT 2323
339 upheValIleThrCysAlaIleGlnMetGlyLysSerGluPheHis 356
    |||||:|||||:|||||
2224 CTTTGTGGCGCGATCTGTCTCATTTGTTTCATATCTTTTGACCAT 2373
356 eHisThrGlnThrThrLeuPheHisThrPheTyraSpLeuLeuIleGln 372
    |||:|||||:|||||:|||||
2374 CCTTGTGATGATGCTGCTGTTCAAGTCTTATGAAACGCTTCTCTTA 2423
373 LysAsnLysHisLysHisLysGlyValAlaIleAspPheIleArg.. 388
    :|||:|||||:|||||
2424 AGGAACAA.....GCGACACTGAAATTCACAAAGC 2455
389 ..SerLeuAspHisCysGlyAspLeuAlaLeuGluGlyValPheSerHisL 405
    :|||||:|||||:|||||
2456 AACGTGTCTCCTGCTGTGAGCTGCGCTTGAAGGCTTTTTCATCTT 2505
405 yspPheAspPheGluLeuGlnAspVal.....SerSerValAsnGluAsp 419
    |||||:|||||:|||||
2506 GCCTTGTGATTAATGATGATCTCGCAGAAAGAGGGGTGATGATGAT 2555
420 ValLeuLeuThrThrGlyLeuLeuCysLysThrThrAlaGlnArgPheLy 436
    |||||:|||||:|||||
2556 GAAGATCTAACCATGCTGTATAGCAAAATTTACAGCCAGAGACTAG 2605
436 sProLysTyrlsPhePheHisLysSerPheGlnGluTyrlThrAlaGly 453
    :|||||:|||||:|||||
2606 ACCATCTACCGGTTTTTAAGTCTCGCTCCAGAAATTTCTGCGGGA 2655
453 rgaArgLeuSerSerLeuLeuThrSerHisGluProGluGluValThrLys 469
    |||||:|||||:|||||
2656 TGAGGCTGATTGAACCTCGATTAGATAGGAGCAAGATCAAGATTG 2705
470 GlyAsnGlyTyrlGluGlnLysMetValSerIleSerAspIleThrSerTh 486
    |||:|||||:|||||
2706 GGACTGATCTATTGAACAACATCACTACCCATGATGACTGTAAAGCG 2755
486 rTyrsSerSerLeuLeuArgTyrlThrCysGlySerSerValGluIaThr 503
    :|||||:|||||:|||||
2756 CTACACAAATTTTGAACATATGTC.....TCCAGCTCCCTTCACAA 2799
503 rGala.....ValMetLysHisLeuAlaIleValTyrlGlnHisGly 516
    :|||:|||||:|||||
2800 AAGCAGGCGCCCAAAATTTGTCTCATTTGCTCCATTAGTGTGATAC... 2846
517 CysLeuLeuGlyLeuSerIleAlaLysArgProLeuThrArgGlnGluSe 533
    :|||:|||||:|||||
2847 .....AAAGAGTC 2854
533 rLeuGlnSerValLysAsn..... 539
    :|||||:|||||:|||||
2855 ATTGGAAGATATATCTGAAGATGATGACTTAAGACACAGCAGAA 2904
540 ..ThrThrGluGlnGluIleLeuLysAlaIle.....AsnIle 551

```

```

2905 TTTCACCTCAAGATGACATTAAGGCGATTGTGCAAAATTTGCCACAA 2954
    :|||:|||||:|||||
552 AsnSerPheValGluCysGlyIleHisLeu..... 561
2955 GCTTACTTTCAATGTTTCAGAACATTTACTGTTCTTGCCCTGAAAC 3004
562 ....TyrlGlnLysSerThrSerLysSerAlaLeuSerGlnGluPheGlu 577
    |||||:|||||:|||||
3005 TGCTTATCAAGCAAACT...GTTGCTGCGTGTCTCATTGTTTTC 3051
577 larpPheGlnGlyLysSerLeuTyrlIleAsnSerGlyAsnIleProAsp 593
    :|||||:|||||:|||||
3052 AATTCCTTCAAGGAGAACACTGACTTGGTGCGCTTAACCTTA...CAG 3098
594 TyrlLeuPheAsp..... 597
3099 TACTTTTGCACACCCAGAAAGCTTGTATGTTGAGAGCATTCACCTT 3148
598 .....PhePheGluHisLeuP 603
3149 CTCATATACGGAATAAGACATCACCCAGACACATTTTTCAGTCTGG 3198
603 roAsnCys.....AlaSerAlaLeuAspPheIleLysLeuAspPheTy 617
    :|||||:|||||:|||||
3199 AAACATGTTTTCAGAAATACACAGGTGCCACTATATATCAGACTATGCT 3248
618 GlyGly.....AlaMetAlaSerTyrGlu..... 625
3249 TCTGCTTTGAACTATGATGAATGGAGGCAAAATTTAGCTGAAAGAA 3298
626 .....LysAlaIleGluAspT 631
3299 GGATATGCTAAAGACATATATGATATGATACGACGAGGACATCACACGC 3348
631 hrGlyGlyIleHisMetClnGluIaIleProGluThrTyrl..IleProSer 646
    :|||:|||||:|||||
3349 TTAAGTACTGCTATGGAACCTTCTCCAAAGCAGTACAAAGATTC... 3395
647 ArgAlaValSerLeuPhePheAsnTyrLysGlnGluPheArgTyrlLeuG1 663
    |||||:|||||:|||||
3396 .....TGTCTAGA 3403
663 uValThrLeuArgAspPheSerLysLeuAsnLysGlnAspIleThrTyrl 680
    :|||||:|||||:|||||
3404 AGTCGATCTCAATATATGATGTTGTAGCCAGAGATCTGTGAGATTC 3453
680 euGlyLysIlePheSerSerAlaThrSerLeuArgLeuGlnIleLysArg 696
    :|||||:|||||:|||||
3454 TAAATGACAGTTTCTCAGCTTCACAGCGCATCGAATCTCATTTAAACAC 3503
697 CysAlaGlyValAlaGlySerLeuSerLeuValLeuSerThrCysLys.. 712
    |||:|||||:|||||
3504 AGCAGAGCGCTTATAGAAGCATCCGCCAGCTCTTAGCTGTCTAAGC 3553
713 AsnIleTyrsSerLeuMetValGluAlaSerProLeuThrIleGluAspG 729
    :|||||:|||||:|||||
3554 CTGTGTCACCAAGTCCTCATAAGCAAGTTGGAATCTGAGCCAGCGAAC 3603
729 luaArgHisIleThrSerValThrAsnLeuLysThrLeuSerIleHis... 744
    :|||||:|||||:|||||
3604 AGGAACGCTTCTCAGCCCTGCTCCCTCGGAATCTTGAAGTCTCAGG 3653
745 .....AspLeuGlnAsnGlnArgLeuProGlyLeuThrAspSerLe 759
    :|||||:|||||:|||||
3654 ACAATCCAGTCAAGACCAATATCTTCTTAAT.....CTGATTAAGTT 3697
759 uGlyAsnLeuLysAsnLeuThrLysLeuIleMetAspAsnIleLysMet 776
    :|||||:|||||:|||||
3698 CCTGTGCTGAAGAACGTCTGTGATCTGAGGCGCATTAATATGTTT 3747
776 snGluGluAspAlaIleLysLeuAlaGluGlyLysAsnLeuLysLys 792
    :|||:|||||:|||||

```

[illegible]

THIS PAGE BLANK (USPTO)

OM of: US-09-697-089-2 to: EST: * out_format: pfs
Date: Mar 26, 2002 2:05 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=fmet+p2n.model -DEV=xih
-O/cgnt2_1/USPFO.spool/US09697089/runatc_25032002_105137_11138/app_query.fasta_1.1097
-DB=EST -QMT=fastap -SUFFIX=est -GAPOP=12.000 -GAPEXT=4.500
-MIMAMCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000 -GAPOP=4.500
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=bloms62 -TRANS=human40.ccl
-LIST=45 -DOCALLIN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500
-MINLEN=0 -MAXLEN=200000000 -USER=US09697089_ECGN1_1.4971
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLBPX -WAIT -THREADS=1

Search information block:

Query: US-09-697-089-2
Query length: 1024
Database: EST: *
Database sequences: 11351937
Database length: 107921985
Search time (sec): 1177.210000

score list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
gb_gss:AK0309404	-	952.00	1.3e-91	552	! A0309404 CTRBI-E1-2528J13.TF CT
gb_gss:AK2360053	-	811.00	1.9e-76	602	! A2360053 IM0103H11F Mouse 10kb
gb_est1:AV1919179	-	801.00	1.5e-75	480	! AV1919179 AV719179 GLC Homo sapi
gb_est1:AV656315	-	593.00	1.2e-53	371	! AV656315 AV656315 GLC Homo sapi
gb_est1:AI2630324	-	585.00	1.5e-52	364	! AI2630324 RST29913 Athieris RAGE
gb_est2:BG210375	-	475.50	1.2e-40	476	! BG210375 RST29913 Athieris RAGE
gb_gss:AK0624020	-	445.50	8.3e-37	470	! A0624020 HS_53F8_B2.C12_Sp6e RH
gb_est1:AW337918	-	377.00	7.1e-38	261	! AW337918 he12h11.x1 NCI.CGAP_CM
gb_gss:AK089169	-	323.00	6.0e-34	404	! A089169 HS_2161_B1.A01.T7C CIT
gb_gss:AK030928	-	298.00	1.5e-21	553	! A030928 RST29913 Athieris RAGE
gb_gss:AK028386	-	260.00	4.8e-22	219	! A028386 RST29913 Athieris RAGE
gb_gss:AK0112439	-	255.00	5.0e-18	553	! A0112439 RST29913 Athieris RAGE
gb_est2:BF903662	-	248.50	4.6e-17	630	! BF903662 CIT-HSP-2372C1.TF CIT
gb_gss:AK2614134	-	244.00	4.4e-16	625	! A2614134 IM0442M17R Mouse 10kb
gb_est1:AW418826	-	222.00	4.0e-12	509	! AW418826 ha21e11.x1 NCI.CGAP_K4
gb_est2:BF207840	-	205.50	3.6e-19	840	! BF207840 601862546F1 NIH.MGC_53
gb_gss:AK2720059	-	189.50	3.3e-93	728	! A2720059 RST29913 Athieris RAGE
gb_est2:HA25984	-	174.00	3.1e-14	546	! A25984 Y156G07.r1 Soares breast
gb_est1:AI023795	-	174.00	3.1e-14	412	! A023795 ox08403.x1 Soares breast
gb_gss:AK2805292	-	171.50	3.1e-12	499	! A2805292 2M0066B12R Mouse 10kb
gb_hlc:AK017656	-	163.00	2.8e-07	1688	! AK017656 Mus musculus 8 days e
gb_hlc:AK017780	-	163.00	2.8e-07	1688	! AK017780 Mus musculus 8 days e
gb_gss:AK019545	-	156.00	2.5e-15	3750	! AK019545 Mus musculus adult ma
gb_gss:AK2762115	-	143.50	2.3e-37	668	! A2762115 IM0556N1R Mouse 10kb
gb_hlc:AK015024	-	143.00	2.3e-35	2043	! AK015024 Mus musculus adult ma
gb_hlc:AK009342	-	141.50	2.4e-21	1324	! AK009342 Mus musculus adult ma
gb_hlc:AK016782	-	141.50	2.4e-21	1324	! AK016782 Mus musculus adult ma
gb_est2:BG448903	-	141.00	2.4e-25	0.0001	! BG448903 NF035641R Mus musculus
gb_est1:AU132668	-	139.50	2.4e-20	861	! AU132668 NF035641R Mus musculus
gb_hlc:AK019775	-	139.00	2.3e-53	2324	! AK019775 Mus musculus adult ma
gb_hlc:AK017814	-	138.00	2.3e-06	2026	! AK017814 Mus musculus 8 days e
gb_hlc:AK011726	-	138.00	2.3e-06	2026	! AK011726 Mus musculus 10 days e
gb_gss:AK0967565	-	136.50	2.37.87	737	! AK0967565 LERIS917F LERG Arabid
gb_est1:AL138332	-	136.00	2.37.59	691	! AL138332 DKEF762N043.r1 762 (S
gb_hlc:AK017842	-	135.00	2.26.17	0.0020	! AK017842 Mus musculus 8 days e
gb_gss:AK2547972	-	134.00	2.23.54	851	! A2547972 EMTD777F Entamoeba h
gb_hlc:AF054176	-	133.00	2.19.48	2218	! AF054176 Homo sapiens angioten
gb_est2:BE83335	-	132.00	2.23.11	384	! BE83335 OV3-OT0055-280600-250-
gb_hlc:AK014400	-	132.00	0.0132	4013	! AK014400 Mus musculus adult ma

gb_est2:BF499764 + 131.00 228.40 0.0015 664 ! BF499764 AT14482.5prime AT D
gb_est2:BG684008 + 130.50 225.40 0.0022 805 ! BG684008 602635613F1 NIH.MGC
gb_est2:BI409867 + 130.00 222.34 0.0032 981 ! BI409867 602962204F1 NCI_CGA
gb_gss:CNS02NV1 - 129.50 220.78 0.0039 1039 ! AL205894 Tetraodon nigrovir
gb_est2:BG994334 - 129.00 225.87 0.0020 586 ! BG994334 PMO-HT1166-130201-0

seq_name: gb_gss:AK0309404

seq_documentation_block:

LOCUS AK0309404 552 bp DNA 22-DEC-1998

DEFINITION CTRBI-E1-2528J13.TF CTRBI-E1 Homo sapiens genomic clone 2528J13,
DNA sequence.

ACCESSION AK0309404
VERSION AK0309404.1 GI:4041438

KEYWORDS GSS.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 552)
Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
Venter,J.C.

AUTHORS Use of a random human BAC End Sequence Database for Sequence-Ready
Map Building

TITLE Unpublished (1998)
Other_GSSs: CTRBI-E1-2528J13.TF

COMMENT Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hncetlgr.org
Clones are available from Research Genetics (info@resgen.com). BAC
and search page:
http://www.figr.org/tbdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.

FEATURES Location/Qualifiers
1..552
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2528J13"
/clone_1="CTRBI-E1"
/sex="male"
/cell_type="sperm"
/note="Vector: pBeloBAC11, Site_1: EcoRI, Site_2: EcoRI;
Caltech Human BAC Library D"

BASE COUNT 144 a 121 c 130 g 157 t
ORIGIN

alignment_scores: 952.00 Length: 183
Ratio: 5.202 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AK0309404/rev ..

Align seg 1/1 to reverse of: AK0309404 from: 1 to: 552

477 MetValSerIleSerAspIleThrSerThrTySerSerLeuLeuArgTty 493
|||||
550 AAGGTTTCATTCGACATTCATTCATTCATTCATTCATTCATTCATTCGGA 501

493 TThrCysGlySerSerValGluAlaThrArgAlaValMetCysHisLeuA 510
|||||
500 CACCTGTGGGTATCTGTGGAAGCACAGGCTGTATGAAGCACCTGG 451

510 LAlaValTyrGlnHisGlyCysLeuGlyLeuSerIleAlaTyrArg 526
|||||

```

450 CAGCAGTGTATCAACACGCGCTCTCGACTTTCATCCAGCAGAG 401
527 ProLeuTPargIngluSerLeuGlnSerValIysasnThrThrGlnG 543
|||||
400 CGCTCTGGAGACAGGAACTTTGCAAAAGTGTAACCAACCACTGAGCA 351
543 ngIuIleuLysAlaIleasnIleasnSerPheValGluCysGlyIleH 560
|||||
350 AGAAATCTCAAGCCATTAACATCAATTCCTTTGTAGAGTGGCATCC 301
560 ILeuTYrGlnGlnSerThrSerIleuSerIleuSerGlnGluPheGlu 576
|||||
300 ATTATATTCAGAGAGTACATCCAAATGAGCCCTGAGCCCAAGATTGAA 251
577 AlaPhePheGlnGlySerSerLeuTYrIleasnSerGlyAsnIleProAs 593
|||||
250 GCATTCTTTCAGGTAAAGCTTATATATCACTGACGAGCAATCCCCGA 201
593 pTYrLeuPheAspPhePheGlnIleuSerProAsnCysAlaSerAlaLeuA 610
|||||
200 TTACTTATTTGACTCTTTGACATTTGCCCAATTTGCAAGTCTCTCG 151
610 sPheIleuLysLeuAspPheTYrGlyGlyAlaMetAlaSerTPGJulys 626
|||||
150 ACTTCATTTAACTGAGCTTTATGAGGAGCTATGGCTTATGGGAAAG 101
627 AlaAlaGluAspThrGlyGlyIleHisMetGluGluAlaProGluThrTY 643
|||||
100 GCTGAGAAACACACAGGTGGAATCCACATGGAAGAGCCCAAGAACTTA 51
643 rIleProSerArgAlaValSerLeuPheAsnTPYrGlnGluPhe 659
|||||
50 CATTCACGACGAGGCTGTATCTTTGTTCTTCACATGGAACAGGAATTC 2
seq_name: gb_gss:AZ360053

```

seq_documentation_block:

```

LOCUS      AZ360053      602 bp      DNA      02-OCT-2000
DEFINITION 1M0103H11F Mouse 10kb plasmid UGCC1M library Mus musculus genomic
clone UGCC1M0103H11 F, DNA sequence.
ACCESSION  AZ360053
VERSION     AZ360053.1 GI:10473753
KEYWORDS
SOURCE      house mouse.
ORGANISM    Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 602)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tinney,A., von Niederhausen,A.
and Wright,D.,Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts

```

TITLE published (2000)

JOURNAL CONTACT: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert length: 10000 Std Error: 0.00
Plate: 0103 row: H column: 11
Seq primer: GGTGTAAACGACGCGCCACT
Class: plasmid ends
High quality sequence stop: 602.
Location/Qualifiers
1..602
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"

FEATURES

1..602
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"

```

/clone="U0GC1M0103H11"
/clone_11b="Mouse 10kb plasmid UGCC1M library"
/sex="Male"
/lab_host="E. coli strain XL10-Gold, Tl-resistant, F-"
/note="Vector: pMD22my. Purified genomic DNA from M.
musculus C57BL/6J (male). Was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using prepared agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD22 (g1473211419b1AE129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT      139 a      153 c      139 g      171 t
ORIGIN

```

alignment_scores:

```

Quality: 811.00      Length: 195
Ratio: 4.481      Gaps: 0
Percent Similarity: 92.821      Percent Identity: 79.487

```

alignment_block:

US-09-697-089-2 x AZ360053/rev ..

Align seg 1/1 to reverse of: AZ360053 from: 1 to: 602

```

382 ALAAlaSerAspPheIleArgSerLeuAspHisCysGlyAspLeuAla 398
|||||
587 GCTTCAGGTATTTTGCACAGAGCGCTGACTGATGAGACACGCGCCCT 538
398 uGluGlyValPheSerHisIysPheAspPheGluLeuGlnAspValSer 415
|||||
537 AGAAGGTGTCTGCCCAAAATTTGATTTGAACCCGACCATGGGTCA 488
415 eValaGlnuAspValIleuLeuThrGlyLeuLeuCysIysTYrThr 431
|||||
487 GCATGAACGAGAGCTCTGTCGACATAGGCGCTCTGTAACTACACA 438
432 AlaGlnArgPheIysProLYrTYrLysPhePheHisIysSerPheGln 448
|||||
437 GCTCAGAGCGCTGAAGCCACGATTAATTTCTTCATTAATCATTTACAGA 388
448 uTYrThrAlaGlyArgArgLeuSerSerLeuLeuThrSerHisGluProG 465
|||||
387 GTACACGAGGCTGCGAGACTCAGCAGTTTGTCTGATCCCAAAAGACAG 338
465 IuGluValThrLysGlyAsnGlyTYrLeuGlnIysMetValSerIleSer 481
|||||
337 AGGAGGTGACCAAGGAGACACTTAACAAATGGTTTCATCTCT 288
482 AspIleThrSerThrTYrSerSerLeuLeuArgTYrThrCysGlySerSe 498
|||||
287 GACATCAATCCCTATATGCAATCTCTCTCAGACGAGTGGGTGTC 238
498 rValGlnAlaThrArgAlaValMetLysHisLeuAlaIleValTYrGlnH 515
|||||
237 CACGAGCAACCAACGAGGCGGTCAAGAGCACCTTGCAATGTTATACGC 188
515 IsgIyCysLeuLeuGlyLeuSerIleAlaIysArgProLeuTPArgIn 531
|||||
187 ACGGACCTTACAAAGACCTTTCAGTACCAAGAGGCTCTCTGAGGACAG 138

```

```

532 Glus2r1euglnserVallysasnrrtrrhglnclglnleu1leuysal 548
|||||.....|||||.....
137 GATCATCATGACAGTGTGAGAAATACACGACGACAGATGATTTGAAAGC 88
548 attaasn1leasnSerPheValGluCysGly1leht5leuTyGlnGluS 565
|||||.....|||||.....
87 CATCATGTGTAATTCCTTCGTAGAGTGTGGCATCATTTCTTCTCAGACA 38
37 GTATGTCTAAATCAGACACTGAGCCAGCAAGATTTGAA 3

seq_name: gb_est1:AV719179

seq_documentation_block:
LOCUS      AV719179      480 bp      mRNA      EST      16-OCT-2000
DEFINITION AV719179 GLC Homo sapiens cDNA clone GLCEQA10 5', mRNA sequence.
ACCESSION  AV719179
VERSION     AV719179.1 GI:10816331
KEYWORDS    EST.
ORIGIN      human.
ORIGIN      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 480)
AUTHORS     Qian,B., Wu,T., Huang,O., Huang,C., Kang,B., Gao,X., Xu,Z., Xiao,H.,
            Xu,X., Li,N., Peng,Y., Liu,F., Qu,J., Song,H., Cheng,Z., Zeng,L.,
            Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Yang
            ,Y., Gu,Y., Chen,Z. and Han,Z.
            Homo sapiens cDNA GLC clones
            Unpublished (2000)
            Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ext. 45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES
source      Location/Qualifiers
            1..480
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="GLCEQA10"
            /clone_id="GLC"
            /tissue_type="corresponding non cancerous liver tissue"
            /dev_stage="Adult"
            /lab_host="SOLR"
            /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
            XhoI"
BASE COUNT      138 a      120 c      85 g      137 t
ORIGIN

alignment_scores:
Quality:      801.00      Length:      156
Ratio:        5.135      Gaps:      0
Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AV719179/rev ..

Align seg 1/1 to reverse of: AV719179 from: 1 to: 480

869 Htsglu1eul1leasparGmetasnVal1leuGln1leuThrlaleu 885
|||||.....|||||.....
480 CATGACATGATCACAGATGAGACGTGCTGAGACACCTCACCCACTGAT 431
885 tleuprTprGlyCysAspValGlnGlnSerLeuSerLeuSerLeuLysH 902
|||||.....|||||.....
430 GCGTCCCTGGGGCTGTGACGTGCAGACGACGCCGTGACAGCCTTTGAAAC 381
902 tleuGlnGlnValProGln1leuVal1lysleuGln1leuLysasnrrtpary 918

```

seq_name	gb_estl	AV656315
seq_documentation_block		
LOCUS	AV656315	371 bp mRNA
DEFINITION	AV656315	GLC Homo sapiens cDNA clone GICEQA10 3', mRNA sequence.
ACCESSION	AV656315	
VERSION	AV656315.1	GI:9877329
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
AUTHORS	1 (bases 1 to 371)	
	Qian,B., Wu,T., Huang,Q., Huang,C., Kang,B., Gao,X., Xu,Z., Xiao,H., Xu,X., Li,N., Peng,Y., Liu,F., Qu,J., Song,H., Cheng,Z., Qu,J., Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Yang,Y., Gu,Y., Chen,Z. and Han,Z.	
	Homo sapiens cDNA clone	
	Unpublished (2000)	
TITLE	Contact: Zeguang Han	
JOURNAL	Chinese National Human Genome Center at Shanghai	
COMMENT	351 Guo Shoujing Road, Zhongjiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China	
	Tel: 86-21-50801919(ex.45)	
	Fax: 86-21-50801922	
	Email: hanzq@chgc.sh.cn	
	This clone is available at CHGC in Shanghai.	
FEATURES	Location/Qualifiers	
SOURCE	1..371	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/clone="GICEQA10"	
	/clone_id="GLC"	
	/tissue_type="corresponding non cancerous liver tissue"	
	/dev_stage="Adult"	
	/lab_host="SOLR"	
	/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2: XhoI"	
BASE COUNT	112 a 85 c 91 g 82 t 1 others	
ORIGIN		

alignment_scores: Length: 141
 Quality: 593.00 Gaps: 1
 Ratio: 4.861
 Percent Similarity: 86.525 Percent Identity: 86.525

alignment block:
 US-09-697-089-2 x AV656315 ..

Align seg 1/1 to: AV656315 from: 1 to: 371

```

625 GLUYSALAAGLUAAPTHRGLYLLEHISMETGLUGLALPROGL 641
      |||
1  GAAAGGCTGCAGAGACACAGGTGGAATCCACATGGAGAGGCCCGCA 50
641 uThrTyrIleProSerArgAlaValSerLeuPheAsnTrpLysGln 658
      |||
51 AACCTACATTCACAGAGGCGCTGATCTTCTTCAACTGGAGACAGG 100
658 LuPheArgThrLeuGluValThrLeuArgAspPheSerLysLeuAsnLys 674
      |||
101 AATTGACAGACTGTGAGGTACACTCCGGGATTTTCAGCAAGTTGAATAG 150
675 GluAspLleThrTyrLeuGlyLysIlePheSerSerAlaThrSerLeuArg 691
      |||
151 CAGATATTCAGATATCTGGGGAAATATTCAGCTCTGCCACAGCCTCAG 200
691 GLeuGlnIleLysArgCysAlaGlyValAlaGlySerLeuSerLeuVal 708
      |||
201 GCTGCAATATGAAGAGATGTGCTGGTGGCTGGAGCCTCAGTTGGTCC 250
708 euSerThrCysLysAsnIleTyrSerLeuMetValGluAlaSerProLeu 724
      |||
251 TCACACCTGTGAAGACATTTATCTCATGTGTGGAAGCCAGTCCCTC 300
725 ThrIleGluAspGluArgHisIleThrSerValThrAsnLeuLysThrLe 741
      |||
301 ACCATAGAAAGTGAGAGCAGCATCATCTGTAAACAACCTG..... 342
741 uSerIleHisAspLeuGlnAsnGlnArgLeuProGlyGlyLeuThrAsps 758
      |||
343 .....GACA 346
758 eLLeuGlyAsnLeuLysAsnLeu 765
      |||
347 GCTTGGGTAACTGAGAAACCTT 369
```

seq_name: gb_est1:A1263294

documentation block:

US A1263294 364 bp mRNA EST 03-FEB-1999
 DEFINITION qx57b01.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2005417 3',
 mRNA sequence.

ACCESSION A1263294

VERSION A1263294.1 GI:3871497

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 364)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov

Life Technologies catalog #: 11548-013

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www.bio.llnl.gov/bbrp/image/image.html

Insert length: 2146 Std Error: 0.00

Seq primer: -40UP from Gibco

FEATURES High quality sequence stop: 364.
 Location/Qualifiers
 source
 1..364

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2005417"
 /clone_1ib="NCI_CGAP_Pan1"
 /tissue_type="adenocarcinoma"
 /lab_host="DH10B"
 /note="Organ: pancreas; Vector: pCMV-Sport6; Site:1; Salt:
 site_2: NotI; Cloned unidirectionally. Primer: Oligo df.
 Average insert size 1.72 kb. Life Technologies catalog #:
 11548-013"

BASE COUNT 117 a 84 c 55 g 108 t
 ORIGIN

alignment_scores: Length: 114
 Quality: 585.00 Gaps: 0
 Ratio: 5.132
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
 US-09-697-089-2 x A1263294/rev ..

Align seg 1/1 to reverse of: A1263294 from: 1 to: 364

```

911 LysLeuGlyLeuLysAsnTrpArgLeuThrAspThrGluIleArgIleLe 927
      |||
364 AACCTGGGTTGAAAACCTGGAGACTCAGCATACAGAGATTAAATTTT 315
927 uGlyAlaPhePheGlyLysAsnProLeuLysAsnPheGlnIleuAsnL 944
      |||
314 AGTGCAATTTTGGAAAGAACCTCGAAAACCTCGACAGAGTTGAATT 265
944 euAlaGlyAsnArgValSerSerAspGlyTrpLeuAlaPheMetGlyVal 960
      |||
264 TGGGGGAAATCGTGTAGCAGTGTATGATGCTTGCATCATGAGTGA 215
961 PheGluAsnLeuLysGlnLeuValPhePheAspPheSerThrLysGluPh 977
      |||
214 TTTCAGAAATCTTAAGCAATTGATGTTTTCACCTTGTACTTAAGATT 165
977 eLeuProAspProAlaLeuValArgLysLeuSerGlnValLeuSerLysL 994
      |||
164 TCTACCTGATCCAGCATTAAGTACAGAAAACCTTACCAAGTCTTATCC 115
994 euThrPheLeuGlnGluAlaArgLeuValGlyTrpGlnPheAspAsp 1010
      |||
114 TTAACCTTTCTGCAAGAGCTAGGCTGTGGTGGAATTTGATGATGAT 65
1011 AspLeuSerValIleThrGlyAlaPheLysLeuValThrAla 1024
      |||
64 GATCTCAGCTGTTATTACAGTGCCTTTTAACTAATACTACTGCT 23
```

seq_name: gb_est2:BG210375

seq.documentation block:

LOCUS BG210375 476 bp mRNA EST 21-APR-2001

DEFINITION RST9913 Athersys RAGE library Homo sapiens cDNA, mRNA sequence.

ACCESSION BG210375

VERSION BG210375.1 GI:13732062

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 476)

AUTHORS Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
 Cahn,S., Dahl,T., Thornton,M., Ramchandran,R., Whittington,J.,
 Lerner,L., Krishnoc,D., McElligott,K., Clark,S., Mays,R., Smith,E.,
 Veloso,N., Hess,J., Cothren,K., Lo,K., Offenbacher,J., Danzig,J.
 and Ducar,M.

TITLE
Creation of Genome-wide Protein Expression Libraries using Random
Activation of Gene Expression
JOURNAL
Nat. Biotechnol. 19 (5), 440 (2001) In press
COMMENT
Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scain@atersys.com

FEATURES
High quality sequence stop: 360.
Location/Qualifiers

SOURCE

1. /476

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="HT1080"

/note="See 'Creation of Genome-wide Protein Expression
Libraries using Random Activation of Gene Expression',
Nature Biotechnology, in press. Note that even though the
cell type indicated is HT1080, since a random activation
method was used, these sequence tags are not necessarily
expressed in HT1080 under normal circumstances."

BASE COUNT 120 a 99 c 107 g 149 t
ORIGIN

alignment_scores:

Quality: 475.50 Length: 152
Ratio: 3.962 Gaps: 3
Percent Similarity: 78.947 Percent Identity: 69.079

alignment_block:

US-09-697-089-2 x BG210375 ..

Align seg 1/1 to: BG210375 from: 1 to: 476

```

882 ThrAlaLeuMetLeuProTrrpGlyCysAspValGlnGlySerLeuSerse 898
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
10 ACTGCAGTGTGCGTCTTGGCTCACTGCCACCTCCGCTTCCCTGGGATTC 59
898 rLeuLeuLys.....HisLeuGlnGluV 906
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
60 TCTGTCTCTATTGCTCTGTGATCTGAAGTATTCCTCGCCACAC 109
906 aLProGlnLeuValLysLeuGlyLeuLysAsnTrrpArgLeuThrAspThr 922
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
110 CTCCTCCAG.....TAGCTGGGATTACCGCTCCAGTCCGCGCACT..... 147
923 GluIleArgIleLeuGlyAlaPhePheGlyLysAsnProLeuLysAsnPh 939
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
148 .....ATFAGTGCATTCTCTGGAAAGAACCCCTCTGAAAAACTTT 185
939 eGlnGlnLeuAsnLeuAlaGlyAsnArgValSerSerAspGlyTrrpLeuA 956
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
186 CCAGCAGCTGAATTGGCGGAAATCGTGTGAGAGTATGATGTCGCTTG 235
956 lAPheMetGlyAlaPheGluAsnLeuLysGlnLeuValAPhePheAspPhe 972
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
236 CTTTGGTGGTGTATTGAGAACTTTAAGCAATTAAGTGTGTTTGTGCTTT 285
973 SerThrLysGluPheLeuProAspProAlaLeuValArgLysLeuSerG1 989
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
286 AGTACTCAAAAGATTTCTACCTGATCCAGCATTAAGTCAGAAAACTTAGCCA 335
989 nValLeuSerLysLeuThrPheLeuGlnGluAlaArgLeuValGlyTrrpG 1006
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
336 AGTGTATCAAGTTAACTTTCTCTGCAGAAAGCTAGCGCTTGTGGGTGCG 385
1006 lAPheAspAspAspLeuSerValIleThrGlyAlaPheLysLeuVal 1022
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
386 AATTGATATGATGATTCACAGCTATTACAGGTGCTTTTAAACTAGTA 435
1023 ThrAla 1024

```

|||||
436 ACTGCT 441

seq_name: gb_gss:A0624020

seq_documentation_block:

LOCUS A0624020 470 bp DNA GSS 16-JUN-1999
DEFINITION HS.5378.B2.C12.SP6E RPCR-11 Human Male BAC Library Homo sapiens
genomic clone Plate=954 Col=24 Row=F, DNA sequence.

ACCESSION A0624020
VERSION A0624020.1 GI:5086412

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

seq. tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCR-11. For BAC
library availability, please contact Pieter de Jong
(pieterdejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Research Genetics (<http://www.hsc.washington.edu>)
Plate: 954 row: F column: 24
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 470.
Location/Qualifiers

FEATURES
SOURCE

1. /470

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Plate=954 Col=24 Row=F"
/clone_lib="RPCR-11 Human Male BAC Library"
/sex="male"

/note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRII. Size selected DNA was cloned into the
pBAC3.6 vector at EcoRI sites"

BASE COUNT 142 a 101 c 107 g 117 t 3 others
ORIGIN

alignment_scores:

Quality: 445.50 Length: 141
Ratio: 4.455 Gaps: 1
Percent Similarity: 70.922 Percent Identity: 67.376

alignment_block:

US-09-697-089-2 x A0624020 ..

Align seg 1/1 to: A0624020 from: 1 to: 470

```

659 pheargThrLeuGluValThrLeuArgAspPheSerLysLeuAsnLysG1 675
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5 TTCATGACTCTGAGGTCACACTCGGCGATTTCAGCAAGTCAATTAAGCA 54
675 nAspIleThrLysLeuGlyLysIlePheSerSerAlaThrSerLeuArgL 692
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

55 AGATATCATGATATCTGGGGAAAAATATTCACGCTCTGCCACAGCCTAGGC 104
692 eucInilelysarGysAlaGlyValAlaGlySerLeuSerLeuValLeu 708
105 TGCAGATTAAGAGATGCTGCTGTGCTGAGGCCCTCATTGCTGCTC 154
709 SerTCysLysAsnIleTySerLeuMetValGlnAlaSerProLeuTh 725
155 AGCAGCTGTAAAGAACCATATCTCTCATGATGAGAACCCAGTCNCCTCAC 204
725 rllleGluAspGluArgHisIleThrSerValThrAsnLeuLysThrIle 742
205 CATAGAGATGACAGAGCAGCATCATCTGTAAACNACCTGAAACCTTTGA 254
742 eTlleHsAspLeuGlnAsnGlnArgLeuProGlyGlyLeuThrAspSer 758
255 GTATTTCATGACCTACAGATCAACGCGCTGCCG..... 286
759 LeuGlyAsnLeuLysAsnLeuThrLysLeuIleMetAspAsnIleLysMe 775
286 ..... 286
775 tAsnGluGluAspAlaIleLeuLysLeuAlaGlyLeuLysAsnLeuLysL 792
287 .....GGTATTGTAATATTCAGTG 305
792 ysmetCysLeuPheHisLeuThr 799
306 GTTTGCTCTTGTTCCTACTTACA 328

```

seq_name: gb_est1:AW337918

seq_documentation_block: 261 bp mRNA EST 31-JAN-2000
 LOCUS AW337918 he12h11.x1 NCI_CGAP_CML Homo sapiens cDNA clone IMAGE:2918853 3',
 DEFINITION mRNA sequence.

ACCESSION AW337918
 VERSION AW337918.1 GI:6834544
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 261)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index

JOURNAL Unpublished (1997)
 CONTACT: Robert Strausberg, Ph.D.
 EMAIL: cgapbs-r@mail.nih.gov

Tissue procurement: Elisabeth Paletta, Jonathan D. Licht, M.D.,
 Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life
 Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The
 I.M.A.G.E. Consortium DNA Sequencing by: Washington University
 Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
 Seq primer: -40UP from Gibco
 High quality sequence stop: 201.

FEATURES

source
 1..261
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2918853"
 /clone_lib="NCI_CGAP_CML1"
 /tissue_type="myeloid cells, 18 pooled CML cases, BCR/ABL
 rearrangement positive, includes both chronic phase and
 myeloid blast crisis"
 /lab_host="DH10B"
 /note="Organ: whole blood; Vector: pCMV-SPORT6; Site:1;
 Salt: Site:2; Note: Cloned unidirectionally. Primer: "
 Oligo dt. Library constructed by Life Technologies."

BASE COUNT 86 a 54 c 38 g 83 t
 ORIGIN
 alignment_scores:
 Quality: 377.00 Length: 74
 Ratio: 5.095 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000
 alignment block:
 US-09-697-089-2 x AW337918/rev ..

Align seg 1/1 to reverse of: AW337918 from: 1 to: 261

```

951 SerAspGlyTrpLeuAlaPheMetGlyValPheGluAsnLeuLysGlnLe 967
260 AGTATGATGATGGCTTGCGCTTCATGGGTGATTTTGAGAGATTTAAGCAAT 211
967 ValPhePheAspPheSerThrLysGluPheLeuProAspProAlaLeu 984
210 AGTGTTTTGGACTTGTACTAAAGAAATTTCTACCGATCCAGCATTAG 161
984 AlaArgLysLeuSerGlnValLeuSerLysLeuThrPheLeuGlnAla 1000
160 TCAGAAAACTTAGCCAGGTATTCACAAATTACTTTCTGCAGAGAGCT 111
1001 ArgLeuValGlyTrpGlnPheAspAspAspLeuSerValIleThrG1 1017
110 AGGCTTGTTGGTGGGCAATTTGATGATGATGATCTCAGTGTATTACAG 61
1017 ValAlaPheLysLeuValThrAla 1024
60 TCCTTTTAAACTAGTACTGCT 39

```

seq_name: gb_gss:AQ889169

seq_documentation_block: 404 bp DNA GSS 10-NOV-1999

LOCUS AQ889169 HS_2161_B1_A01_TTC CIT Approved Human Genomic Sperm Library D Homo
 DEFINITION sapiens genomic clone Plate=2161 Col=1 Row=B, DNA sequence.

ACCESSION AQ889169
 VERSION AQ889169.1 GI:6345359
 KEYWORDS GSS.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 404)
 AUTHORS Mahatras,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
 Hood,L.

TITLE Sequencing-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589

COMMENT Contact: Mahatras GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887

Email: jwallace@u.washington.edu
 Clones may be purchased from Research Genetics (info@resgen.com).
 BAC end Web Server: <http://www.htsc.washington.edu>
 Plate: 2161 row: B column: 1
 Seq primer: 77

Class: BAC ends
 High quality sequence stop: 404.

FEATURES

source
 1..404
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=2161 Col=1 Row=B"

/clone.lib="CIT Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
 E-Coli DH10B"
 BASE COUNT 119 a 93 c 70 g 122 t
 ORIGIN

alignment_scores:

Quality: 323.00 Length: 90
 Ratio: 3.892 Gaps: 0
 Percent Similarity: 92.222 Percent Identity: 77.778

alignment_block:

US-09-697-089-2 x AQ889169/rev ..

Align seg 1/1 to reverse of: AQ889169 from: 1 to: 404

```

1 MetAnpHeilleLysAspAsnSerArgAlaLeuIleGlnArgMetGlyme 17
   :::::::::::::::::::: ::::::::::::::::::::::::::::::
351 GTCATATTTCATAGACGACGATACCCAGC.CTTATTCAAAAGATGGCAAT 303
   :::::::::::::::::::: ::::::::::::::::::::::::::::::
17 ttrvAlleLysGlnIleThrAspAspleuPheValTTPAsnValleua 34
   :::::::::::::::::::: ::::::::::::::::::::::::::::::
302 GCCTGTATAGACTACTCAGATGACCTATTGTATGAATGTTATGA 253
   :::::::::::::::::::: ::::::::::::::::::::::::::::::
34 snArgLugluValAsnIleIleCysGlnLysValGluGlnAspAla 50
   :::::::::::::::::::: ::::::::::::::::::::::::::::::
252 ATCGAGAGAGAGTAACATCATTTCTGGAGAGGTGGAGCGAGATGCT 203
   :::::::::::::::::::: ::::::::::::::::::::::::::::::
51 AlaArgGlyIleIleHisMetIleLeuLysGlySerGlnSerCysAs 67
   :::::::::::::::::::: ::::::::::::::::::::::::::::::
202 GTTAGAGGATCATTCACATGATTTGMAAGAGGTTCAAGATCCTGTA 153
   :::::::::::::::::::: ::::::::::::::::::::::::::::::
67 nleuPheLysSerLeuLysGluTTPAsnTyProLeuPheGlnAsp 84
   :::::::::::::::::::: ::::::::::::::::::::::::::::::
152 CCTCTATGTTAAATCCCTTAAGGAGTGAACATATCTATTTCAAGACT 103
   :::::::::::::::::::: ::::::::::::::::::::::::::::::
84 euAsnGlyGlnSerLeuPhe 90
   :::::::::::::::::::: ::::::::::::::::::::::::::::::
102 TGAAGGACAAAGTAAGTAT 83

```

seq_name: gb_gss:AQ320928

seq_documentation_block:

LOCUS AQ320928 553 bp DNA GSS 06-MAY-1999
 DEFINITION RPI11-93C9_TV RPI1-11 Homo sapiens genomic clone RPI1-11-93C9, DNA
 sequence.
 ACCESSION AQ320928
 VERSION AQ320928.1 GI:4053662
 KEYWORDS GSS.

SOURCE

ORGANISM human.

REFERENCE

AUTHORS
 TITLE
 JOURNAL
 COMMENT

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 553)
 Adams M.D., Rounsley S.D., Zhao S., Bass S., Linher K., Golden K.,
 Berry K., Granger D., Suh E., Wible C., de Jong P. and Venter J.C.
 Use of human BAC End Sequences for Sequence-Ready Map Building
 Unpublished (1998)
 Other GSSs: RPI11-93C9_TV
 Contact: Shaying Zhao, William Nierman, Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: hbeetlgr.org
 Clones are derived from the human BAC library RPI1-11. For BAC
 library availability, please contact Pieter de Jong
 (pieter@jgong.med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
 Research Genetics (info@resgen.com). BAC end search page:

http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html
 Seq primer: 77
 Class: BAC ends.

FEATURES

source

Location/Qualifiers
 1..553

/organism="Homo sapiens"
 /db_xref="GDB:7535384"
 /db_xref="taxon:9606"
 /clone="RPI1-11-93C9"
 /clone.lib="RPI1-11"
 /sex="Male"
 /cell_type="Lymphocytes"
 /note="Vector: pBAC3.6; Site 1: EcoRI; Site 2: EcoRI;
 RPI11 Human Male BAC library"
 BASE COUNT 170 a 107 c 114 g 162 t
 ORIGIN

alignment_scores:

Quality: 298.00 Length: 74
 Ratio: 4.656 Gaps: 1
 Percent Similarity: 86.486 Percent Identity: 81.081

alignment_block:

US-09-697-089-2 x AQ320928/rev ..

Align seg 1/1 to reverse of: AQ320928 from: 1 to: 553

```

872 lIeAspArgMetAsnValLeuGlnIleuThrAlaLeuMetLeuProtr 888
   :::::::::::::::::::: ::::::::::::::::::::::::::::::
259 GTCGACAGATGAGTAAGCTGCTAAGAAAGCTCACCGCACTGATGCTGCC 210
   :::::::::::::::::::: ::::::::::::::::::::::::::::::
888 pGlyCysAspValGlnGlySerLeuSerSerLeuLeuLysHisLeuGlu 905
   :::::::::::::::::::: ::::::::::::::::::::::::::::::
209 GGCTGTGAGCTGCAAGGACGCTGAGCAGCCGTGGAACATTTGGAG 160
   :::::::::::::::::::: ::::::::::::::::::::::::::::::
905 luvAlProGlnLeuValLysLeuGlyLeuLysAsnTTPArgLeuThrAsp 921
   :::::::::::::::::::: ::::::::::::::::::::::::::::::
159 AGTCCACACACTGCTCAAGCTTGCGTTGAAAAAAGTGAAGATCACACAT 110
   :::::::::::::::::::: ::::::::::::::::::::::::::::::
922 ThrGlnIleArgIleLeuGlyAlaPheNegIlyLysAsnPro..... 935
   :::::::::::::::::::: ::::::::::::::::::::::::::::::
109 ACAGAGATTAGAAATTTAGGTAGTACACACATACAGAGCCAAATTAAT 60
   :::::::::::::::::::: ::::::::::::::::::::::::::::::
936 .....LeuLysAsnPhe 939
   :::::::::::::::::::: ::::::::::::::::::::::::::::::
59 GGATTTGGCCCTTAAATAATTC 38

```

seq_name: gb_gss:AQ283886

seq_documentation_block:

LOCUS AQ283886 219 bp DNA GSS 27-APR-1999
 DEFINITION RPI11-78E13_TV RPI1-11 Homo sapiens genomic clone RPI1-11-78E13,
 DNA sequence.
 ACCESSION AQ283886
 VERSION AQ283886.1 GI:3910204
 KEYWORDS GSS.

SOURCE

ORGANISM human.

REFERENCE

AUTHORS
 TITLE
 JOURNAL
 COMMENT

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 219)
 Adams M.D., Rounsley S.D., Zhao S., Bass S., Linher K., Golden K.,
 Berry K., Granger D., Suh E., Wible C., de Jong P. and Venter J.C.
 Use of human BAC End Sequences for Sequence-Ready Map Building
 Unpublished (1998)
 Contact: Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: mdadams@tigr.org

Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genetics (<http://inforesgen.com>). BAC end search page: http://www.tigr.org/cdb/humgen/bac_end_search/bac_end_search.html
 Seq primer: 17
 Class: BAC ends.

FEATURES

source

1. 219
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="GDB:752676"
 /db_xref="taxon:9606"
 /clone="RPCI-11-78E13"
 /clone_lib="RPCI-11"
 /sex="Male"
 /cell_type="Lymphocytes"
 /note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
 RPCI11 Human Male BAC Library"
 RPCI11 Human Male BAC Library
 69 a 45 c 46 g 59 t

alignment_scores:

Quality: 260.00 Length: 76
 Ratio: 4.194 Gaps: 1
 Percent Similarity: 81.579 Percent Identity: 72.368

alignment_block:

US-09-697-089-2 x AQ283886 ..

Align seg 1/1 to: AQ283886 from: 1 to: 219

```

701 AlaglySerLeuValleuSerThrcysAsnileYrSerle 717
|||||
3 GCTGGAACCTTACTGTGCTCTCAGCACCTGTGAAGACATTATTCCT 52
717 UmetValGluAlaSerProleuThrileGluAspGluArgHisileThrs 734
|||||
53 CATGCTGGAACCCAGTCCCTGACCATGAGATGAGAGCAGCATCATCAT 102
734 erValThrAsnleuSerThrcysAsnileYrSerle 750
|||||
103 GTGTAACAACTGAAACCTTGATTCATGACCTACAGATCAACAGG 152
751 leuProGlyGlyLeuThrAspSerLeuGlyAsnleuYrSerleThry 767
|||||
153 CTGCGCGGT.....ATTGTTATATCAATG 178
767 slenUleMetaspAsnileYrMetasn 776
|||||
179 TGTGTGCTTGTCTCACTTAAAAAAT 206

```

seq_name: gb_gss:AQ112439

seq_documentation_block:

LOCUS AQ112439 630 bp DNA GSS 29-AUG-1998
 DEFINITION CIT-HSP-2372C1.TR CIT-HSP Homo sapiens genomic clone 2372C1, DNA
 sequence.
 ACCESSION AQ112439
 VERSION AQ112439.1 GI:3484599
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 630)
 AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
 Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
 Venter,J.C.
 TITLE Use of a random human BAC End Sequence Database for Sequence-Ready
 Map Building
 JOURNAL Map Building (1998)

COMMENT

Other_GSSs: CIT-HSP-2372C1.TF
 Contact: Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: mdamas@tigr.org
 Clones are available from Research Genetics (<http://inforesgen.com>). BAC
 end search page:
http://www.tigr.org/cdb/humgen/bac_end_search/bac_end_search.html.
 Seq primer: M13 Reverse
 Class: BAC ends.

FEATURES

source

1. 630
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="2372C1"
 /clone_lib="CIT-HSP"
 /sex="Male"
 /cell_type="Sperm"
 /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
 HindIII"
 188 a 126 c 124 g 192 t

BASE COUNT 188 a 126 c 124 g 192 t
 ORIGIN

alignment_scores:

Quality: 255.00 Length: 48
 Ratio: 5.312 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x AQ112439 ..

Align seg 1/1 to: AQ112439 from: 1 to: 630

```

784 AlaglyLeuLysAsnleuLysMetCysleuPheHisleuThri 800
|||||
486 GCTGAAGCCCTGAAAAACCTGAGAGATGTTATTTCATTGACCCA 535
800 sleuSerAspIleGlyGlyMetAspTyrIleVallySerleuThrs 817
|||||
536 CTTCTCTGACATTTGAGAGGAAATGATTACATGTCAGATCTGTCAA 585
817 ergluProCysAspLeuGluIleGlnleuValserCysCys 831
|||||
586 GTGAACCTGTGACCTGACAAATTCATTAGTCTCTGCTGC 629

```

seq_name: gb_est2:BF903662

seq_documentation_block:

LOCUS BF903662 251 bp mRNA EST 18-JAN-2001
 DEFINITION IL2-MT0180-181200-276-F03 MT0180 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BF903662
 VERSION BF903662.1 GI:12295121
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 251)
 AUTHORS Nagai,M.A., da Silva,W. Jr., Zagdo,M.A., Bordin,S., Costa,F.F.,
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
 Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
 M.D., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
 Simpson,A.J.
 TITLE Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 MEDLINE 20202663
 COMMENT Contact: Simpson A.J.G.

Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be found in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?li=IL2&t2=IL2-MT0180-
181200-2/6-F03&t3=2000-12-18&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 45
High quality sequence stop: 96.
Location/Qualifiers

FEATURES

Source

1. 251
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MT0180"
/dev_stage="Adult"
/note="Organ: marrow; Vector: puc18; Site_1: Sma1; Site_2:
Sma1; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the puc 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

BASE COUNT 72 a 62 c 56 g 61 t
ORIGIN

alignment_scores:

Quality: 248.50 Length: 79
Ratio: 3.883 Gaps: 3
Percent Similarity: 81.013 Percent Identity: 74.684

alignment_block:

US-09-697-089-2 x BF903662 ..

Align seg 1/1 to: BF903662 from: 1 to: 251

```

337 ThProLeuPheValIleThrCysAlaIleGlnMetGly_GluSerG 353
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
3 ACCGCTCTTTGAGGTCATCATTTGTGCATCTCATGACAGCAAGG 52
353 lUpHeHisSerHisThrGlnThrThLeuPheHisThrPheTyrAspLeu 369
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
53 AGTTCACCTCTCACACACAACACCTGATCATCTTCTATGATCTG 102
370 LeuIleGlnLysAsnLysHisLysLysGlyValAlaAlaLaserAspH 386
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
103 AGATACAGAAAAACAAACAAACATAAAGGCGTG.....TCTGATGT 146
386 eIleArgSer...LeuAspHisCysGlyAspLeuAlaLeuGluGlyValP 402
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
147 GACTCATTCGAGCGCTGACCATCTGTGG.GACCTAAGATCTGTGGTGGCT 195
402 heSerHisLysPheAspPheGlnLeuGlnAspVal 413
||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
196 TCTGCCACAAAGTTGAATTCGAATCGAATCGACGATGTG 230

```

seq_name: gp_gss:AZ614134

seq_documentation_block:

LOCUS AZ614134 625 bp DNA GSS 13-DEC-2000
DEFINITION IM0442N17R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0442N17 R, DNA sequence.
ACCESSION AZ614134
VERSION AZ614134.1 GI:11736324
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE

1 (bases 1 to 625)
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS

Dunn,D., Aoyagi,A., Barber,M., Beccorn,T., Duval,B., Hamll,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Kelly
,M., Rose,M., Rose,R., Stokes,R., Tinney,A., von Niederhausern,A.
and Wright,D.,Weiss,R.

TITLE

Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts

JOURNAL

Unpublished (2000)

COMMENT

Contact: Robert B. Weiss
University of Utah
Genome Center

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT
84112, USA

Tel: 801 585 5606
Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0442 row: N column: 17

Seq primer: CACACAGCAACACGCTATGACC

Class: plasmid ends

High quality sequence stop: 625.

FEATURES

Source

1. 625
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="UUGC1M0442N17"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PMD42nrv: Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnarep/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PMD42 (g114732114[gblAF129072.1]), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT

171 a 143 c 144 g 167 t

ORIGIN

alignment_scores:
Quality: 244.00 Length: 188
Ratio: 1.952 Gaps: 4
Percent Similarity: 66.489 Percent Identity: 34.043

alignment_block:

US-09-697-089-2 x AZ614134/rev ..

Align seg 1/1 to reverse of: AZ614134 from: 1 to: 625

```

325 lIleGlnLysSerArgCysLeuArgAsnLeuMetLysThrProLeuPheYa 341
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
617 TTGGACAGATGAAGATTTCACAGGAATTCACAAGACCCCTCTTCGT 568
341 ValIleThrCysAlaIleGlnMetGlyGluSerGluPheHisSerHisT 358
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
567 GGCAGCAGTATGTACTGACTGTGTGAAATTCATCTGACACACCCCTTC 518
358 hrcLInThrThLeuPheHisThrPheTyrAspLeuLeuIleGlnLysAsn 374

```

```

517 AGGATATGGACACTTTTCAGTCCGTACATGCATAATCATCTG..... 480
      ::::: ||||| ::::: ::: |||
375 LYSHTLSHSLYSGLYALALAASERASPHEILARGSERLEUAS 391
      ::::: ||||||| ||||| ::::: :::::
479 TCCATTAAACCACAAGCT...CGACGTAAACCCCTCAAGCCACCCTGCC 433
      ::::: ||||||| ||||| :::::
391 RHISCSGYLSPLEUALALEUCILYVALPHERNHSLSYPHESP 408
      ::: ||||||| ||||| :::::
432 CTCATGTGGCAGCTGGCTTGACAGGCTTTCTCATGATGCTTTGAGI 383
      ::: ||||||| ||||| :::::
408 HEGILDEULINAPVAL.....SERSERVALSNGILUSPALLEU 422
      ::: ||||||| ::::: |||||
382 TCAAATAGTAGTACCCGCGCAGAGCGAGGATTCATGAAGAAGCTC 333
      ::: ||||||| ||||| :::::
423 TTHTHGLYLEUCUSLYSTYTPRRALAGNARGPHELYSPROFYSTY 439
      ||||| ||||| ::::: ||||||| ::::: |||
332 ACCACCTGCTTATGAGCAAAATTCACCGCCCAAGACTGAGCCACTGTA 283
      ||||| ||||| ::::: ||||||| ::::: |||
439 RLSPRPHENHSLSSERPHEINGLUYTRHALAGLAYATGATLEUAS 456
      ||::: ||||| ::::: ||||| ::::: |||||
282 CCGGTTTTAGTCCGCTGTCAGAGAGTTCTTGCTGCATGAGACTGA 233
      ::::: ||||| ::::: |||||
456 ERSERLEUTHRSERNHSLGNPROGLUGLVALTHIRYSELVANGLY 472
      ::::: ||||| ::::: |||||
232 CTGAACCTTGAGTTCCATAGGCAAGAACCCACAGATCTGCACCTTAT 183
      ::::: ||||| ::::: |||||
473 TYULEUGILYSMETVALSERILSESARPILETHERSETHYTRISE 489
      ||||| ::::: |||
182 TATTITAGCAAAATTAATCACCCCTGGAAGGCTTACACACCTACACA 133
      ||||| ::::: |||
489 RLEULEAUGTYRTHCYSGLY...SERSERYALGIJALATHARGALAY 505
      ::::: ||||| ::::: |||
132 TTTTTHGAGATATGCTCTTAGCCATCCGTCATCAAAGCAGGCGCAAC 83
      ::::: ||||| ::::: |||
505 ALMETLYSHILEU 509
      ||::: |||||
82 TTGTATTCATCTTG 69

```

```

FEATURES
source      Location/Qualifiers
1. 509      /organism="Homo sapiens"
            /db_xref="taxon:9606"

```

```

/cclone="IMAGE:2874380"
/cclone_lib="NCI_CGAP_Kid12"
/tissue_type="2 pooled tumors (clear cell type)"
/lab_host="DH10B"
/note="Organ: Kidney; Vector: p77T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP_Kids was
prepared, and ss circles were made in vitro. Following NRP
hybridization, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneids 132912-1325831, 1471368-1472903 and
1492104-1493255). Subtraction by Bento Soares and M.
Fatima Ronaldo."
BASE COUNT      161 a      99 c      80 g      169 t
ORIGIN

alignment_scores:
      Quality: 222.00      Length: 45
      Ratio: 4.933      Gaps: 0
Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AW418826/rev ..

Align seg 1/1 to reverse of: AW418826 from: 1 to: 509

980 AspProAlaLeuValArgLysLeuSerGlnValIleuSerLysLeuThrph 996
|||||
508 GATCCAGCATTTAGTGTGCAAACTTACCACAAGTTTATCCAAAGTTAACTTT 459
|||||
996 eLeuGInGluAlaArgLeuValGlyTrpGlnPheAspAspAspLeuS 1013
|||||
458 TCTGCAAGAAGCTAGGCTTGTGGGTGGCAATTGATGATGATGATCTCA 409
|||||
1013 eValIleThrGlyAlaPheLysLeuValThrAla 1024
|||||
408 GTGTTATTACAGGTCTTTAACTGATAGTACTGCT 374

```

BASE COUNT	161 a	99 c	80 g	169 t
ORIGIN				

```
alignment_scores:
  Quality: 222.00      Length: 45
  Ratio: 4.933         Gaps: 0
Percent Similarity: 100.000  Percent Identity: 100.0000
```

```
alignment_block:
```

OS-09-697-089-2 x AW418826/rev

Align seg 1/1 to reverse of: AW418826 from: 1 to: 509

980 *AspProAlaLeuValAlaGlySerLeuSerGlnValLeuSerLysLeuThr* 996
508 *GATCCAGCATTTAGTCAGAAACTTGGCCAGGTATATCCAACTTAACCTT* 459
996 *eleuengingluAaArgLeuValGlyTrpGlnPheAspAspAspLeu* 1011
458 *TCTCCAGACGCTAGCGCTGTGGGTGGCAATTGATCATGATGATCTCA* 409
1013 *erValIleThrGlyAlaPheLysLeuValThrAla* 1024
408 *GTTGATTATACAGTCTCTTTAACTAGTAACTGCT* 374

THIS PAGE BLANK (uspto)


```
score_list:
```

seq_name: qb_pr:AY032589

seq_documentation_block:

LOCUS AY032589 3133 bp mRNA PRI 25-MAY-2001

DEFINITION Homo sapiens caspase recruitment domain protein 12 mRNA, complete cds.

ACCESSION AY032589

VERSION AY032589.1 GI:13899172

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 3133)
Geddes, B. J., Wang, L., Huang, W. J., Lavelle, M., Manji, G. A., Brown, M., Juman, M., Morganstern, J., Merriam, S., Glucksmann, A., Distefano, P. S. and Bertin, J.
Human CARD12 Is a Novel CED4/Apaf-1 Family Member That Induces Apoptosis
Biochem. Biophys. Res. Commun. 284 (1), 77-82 (2001).

TITLE Human CARD12 Is a Novel CED4/Apaf-1 Family Member That Induces Apoptosis

JOURNAL Biochem. Biophys. Res. Commun. 284 (1), 77-82 (2001).

PUBMED 11374873

REFERENCE 2 (bases 1 to 3133)
Bertin, J.
Direct Submission
Submitted (15-APR-2001) Neurobiology, Millennium-Pharmaceuticals Inc., 640 Memorial Drive, Cambridge, MA 02439, USA

AUTHORS

TITLE

JOURNAL

FEATURES

source

1..3133
Location/Qualifiers

/organism="Homo sapiens"

/db_xref="taxon:9606"

36..3110
/note="CARD12: CED4/Apaf-1 family member"

/codon_start=1

/product="caspase recruitment domain protein 12"

/protein_id="AA38730.1"

/db_xref="GI:13899173"

/translation="MNFKDNSRALIQRMGVYIKQITDDLFWYVNLNREEVNIICE
KVEDARAGIIMHLKKGSSCNFLSKEMWYPLFDLNGQSLFQTSSEGDIDIA
QDSKLTIESGSPFNFYLGEDIDIFNLKSTFERYLRKDQNNHRYEOLITLGLDIA
LOSCEITGSESGKSKRLRIORIMALMGSSCKALTKRPVFLRLSAGQGLEETCD
OLIDIPCTIKOPFMMLIKRLRVFLIDGQVFNKQNPCEALTEIKHRRKNMYI
VTTTDECLRHTRIQFGALTAEGDTEDSOALRLVQLKIKELAGSLIQIKOSKRL
MKPLPFLVITCAIQMGSEFHSFHSHTQTLFFPDLQIKKHNKGVAAADFLRSIDH
CGDIALGVPFSHKRDFELQDVSYNEDVLTGLCYTAQRKPKPKFHKSFQEYT
AGRLSSILTSHEPEVTKNGYLOKKWASIDISTYSTSLRLRYTSGSVATAVNKH
LAAYVHOGCLGLSIAKRLPMROSIOSTVNTPEOELKAININSFVEGCIHLYOEST
SKNLSOEPERAFQGSRLYNSGNIIPYLFDFPHPCASADDFIKLDPGNGAMSM
EKAEDDGGIIMEAPETVTPSRVSLFPMWKEPFLVETLVDPSKLNQDITVLTAK
IFSSATSLRLQIKKACAGVASSLSVSTGCNITSLSLWESAPLITTEBRHLLTVNLKT
LSLHIDNQLRIQGLTIDSLGNLNLTKLINDNTKMNEDPAIKTAEKLNKSKLEPL
THSIDLEGMDYIYKSLSEPCDLEETQVSSCLSAVAVKLILKQNLNLVAKLSILDS
ENYLEKDGNAELHLLIDRMVNLQETLALIMPQGDVGSGSLIKLNLHEEVPYKIGL
KNRRLPTETRLRIICGFRKNPKLNFQOLNLACNRVSSDGLAMGVENIKQLVYDFE
STKEFLPDPALVRKLSVLSKTLFLOEARLVGQFDDDDLSVITGAFLVTA"

36..299
/note="CARD: Region: caspase recruitment domain"

522..1406
/note="NBS: Region: nucleotide-binding site"

2001..3107
/note="LRR: Region: leucine-rich repeats"

BASE COUNT 903 a 691 c 729 g 810 t

ORIGIN

alignment_scores:
Quality: 5281.00 Length: 1024
Ratio: 5.157 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x AY032589 ..

Align seg 1/1 to: AY032589 from: 1 to: 3133

```

1 MetAspPheIleLysAspAsnSerArgAlaLeuIleGlnArgMetGlyme 17
|||||
36 ATGAATTCATTAAGACAAATACCGAGCCCTTATTCAAAGATGGGAAT 85
17 ThrValIleLysGlnIleThrAspAspLeuPheValITrpAsnValLeuA 34
|||||
86 GACTGTTATAAAGCAAAATCACAGATGACCTATTGTATGGAATGTTCTGA 135
34 snArgGluGluValAsnIleIleCysGlyLysValGluValGluAsnAla 50
|||||
136 ATCCGCAAGAACTAAACATCATTTGCTGCGAGAGGTGGAGCAGATGCT 185
51 AlaArgGlyIleIleHisMetIleLeuLysGlySerGluSerCysAs 67
|||||
186 CCTAAGGCGATCATTCACATGATTTTGAAGAGGTTTCAGAGTCTGTAA 235
67 IleuPheLeuLysSerLeuLysGluITrpAsnITrpProLeuPheGlnAsp 84
|||||
236 CCTCTTTCTTAAATCCCTTAAGAGAGTGAACATATCTTATTTTCAGAG 285
84 euAsnGlyInSerLeuPheHisGlnThrSerGluLysAspLeuAspAs 100
|||||
286 TGAATGAGCAAAAGTCTTTTCATCAGACATCAGAAAGAGACTTGGACGAT 335
101 LeuAlaGlnAspLeuLysAspLeuITrHisITrProSerPheLeuAsn 117
|||||
336 TTGGCTCAGAGATTAAAGGACTTGTACATACCCCATCTTTCTGAACTT 385
117 eITrProLeuGlyGluAspIleAspIleIlePheAsnLeuLysSerThr 134
|||||
386 TTATCCCTTGCTGAAAGATATGACATATTTTAACTTGAAGAGCACT 435
134 heITrGluProValLeuITrPArgLysAspGlnHisHisITrArgValG 150
|||||
436 TCACAGAACTGTCCTGCTGAGGAAGAGACCAACCATCACCCTGCGAG 485
151 GlnLeuITrLeuAsnGlyLeuLeuGlnAlaLeuGlnSerProCysIle 167
|||||
486 CAGCTGACCTGAAATGGCTCTCTGAGGCTTTCAAGACCCCTGCATCAT 535
167 eGluGlyGluSerGlyLysSerITrLeuLeuGlnArgITrIleAla 184
|||||
536 TGAAGGGGAAATCTGGCAAGGCAAGTCCACTGCTGACAGCGCATTTGCC 585
184 eLeuITrPArgLysGlyLysCysLysAlaLeuITrLysPheLysPheVal 200
|||||
586 TGCTCTGGGGCTCCGGAAGTGCAGAGCTCTGACCAAGTTCAAAATTCGTC 635
201 PhePheLeuArgLeuSerArgAlaGlnGlyLysLeuPheGluITrLeu 217
|||||
636 TTCTTCTCTCCGTCAGCAGGAGCCAGGGGTGGACTTTTGAACCTCTG 685
217 sAspGlnLeuAspIleProGlyITrITrLeuArgLysGluITrITrPhe 234
|||||
686 TGATCAACTCTCGAATATACCTGGCAACAATCAGAGGAGACATTCATG 735
234 IAspLeuLeuLysLeuArgGlnArgValLeuPheLeuLeuAspGlyITr 250
|||||
736 CCATGCTGCTGAAGCTCGCGCAGAGGTTCTTTCTTCTTGATGGCTAC 785
251 AsnGluPheLysProGlnAsnCysProGluIleGluAlaLeuIleLysG 267
|||||
786 AATGATATCAAGCCCAAGAACTGCAGAAATCGAAGCCCTGATTAAGGA 835
267 uAsnHisITrPheLysAsnMetValIleValITrITrITrITrGluCys 284
|||||

```

```

836 AAACACCGCCTTCAGAACATGCTCATGCTCACCACTACCTAGTGGCC 885
284 euArgHisITrLeuArgITrPheGlyAlaLeuITrITrAlaGluValGlyAs 300
|||||
886 TGAAGCACAATACCGCAGTTTGGTGGCTGACTGCTAGGTGGGGGATATG 935
301 ThrGluAspSerAlaGlnAlaLeuIleArgGluValLeuIleLysGlu 317
|||||
936 ACAGAAGACAGCCGCCAGGCTCTCATCCGAGAAAGTCTGATCAAGAGCT 985
317 uAlaGluGlyLeuLeuLeuGlnIleGlnLysSerArgCysLeuArgAsn 334
|||||
986 TGCTGAAGCTTGTGCTCCAAATTCAGAAATCCAGAGTCTTGAGGAATC 1035
334 euMetLysITrProLeuPheValIleITrCysAlaIleGlnMetGly 350
|||||
1036 TCATGAAGACCCCTCTCTTGTGCTATCATCTTGTCATTCAGATGGGT 1085
351 GluSerGluPheHisSerHisITrGlnITrITrLeuPheHisITrPhe 367
|||||
1086 GAAAGTAGTTCACACTCTCACACACAAACAGCTGTCATACCTTTCTA 1135
367 rAspLeuLeuIleGlnLysAsnLysHisLysGlyValAlaAla 384
|||||
1136 TGATCTGTTGATACAGAAAAACAAACATAAAGGTGGCTGCA 1185
384 eITrAspPheIleITrArgSerLeuAspHisCysGlyAspLeuAlaLeu 400
|||||
1186 GTACACTTCAITTCGAGACCTGAGACACTGTGAGACCTACTCTGGAGGT 1235
401 ValPheSerHisLysPheAspPheGluLeuGlnAspValSerSerVal 417
|||||
1236 GTGTTCTCCACAAAGTTGATTTGCAATGCAAGATGTCTCAGCGTGA 1285
417 nGluAspValLeuLeuITrITrGlyLeuLeuCysLysITrITrAlaGln 434
|||||
1286 TGAAGATGTCCTGCTGACAACTGGGCTCTCTGTAATATTAAGCTCAA 1335
434 rPheLysProLysITrLysPhePheHisLysSerPheGlnGluITrITr 450
|||||
1336 GGTTCAAGCAAGATATTAATTTCTTTCACAGTCATTCAGGAGTACAA 1385
451 AlAlaGlyArgLeuSerSerLeuLeuITrSerHisGluProGluVal 467
|||||
1386 CGAGGACGAAGACTCAGCACTTATGACGTCATCAGACCCAGAGAGGT 1435
467 ITrITrLysGlyAsnGlyITrLeuGlnLysMetValSerIleSerAsp 484
|||||
1436 GACCAAGGGGAATGTTACTTGCAGAAAAATGGTTTCATTTGCGACATTA 1485
484 hrSerITrITrSerSerLeuLeuArgITrITrCysGlySerSerValGlu 500
|||||
1486 CATCCACTTATAGCAAGCTGCTCCGCTACACTGTGGCTGATCTGGGAA 1535
501 AlITrITrArgAlaValMetLysHisLeuAlaAlaValITrGlnHisGly 517
|||||
1536 GCCACACGAGGCTGTTATGAAGCACCTCGCAGAGTATCAACACGCGCT 1585
517 sLeuLeuGlyLeuSerITrAlaLysArgProLeuITrPArgGlnGlu 534
|||||
1586 CCTCTCGGACTTTCATCGCCAGAGGCTCTCTGAGACAGGAATCTT 1635
534 euGlnSerValLysAsnITrITrGluGlnGluIleLeuLysAlaIleAs 550
|||||
1636 TGCAAAAGTGTGAAGAAACCACTGAGCAAGAAATTTGAAACCAATAAC 1685
551 ITraAsnSerPheValGluCysGlyIleHisLeuITrGlnGluSerITr 567
|||||
1686 ATCAATTCCTTTGTAGAGTGTGGCATCCATTTATATCAAGAGATACAT 1735
567 rLysSerAlaLeuSerGlnGluPheGluAlaPhePheGlnGlySerITr 584
|||||
1736 CAATCAGCCCTGAGCCAAAGAAATTTGAAGCTTTCTTCAAGGTAAGCT 1785

```

584 eutyrtlleasnserylYasnilleproaspyrilleupheasphepGlu 600
 1786 TATATATCAACTCAGGAGAACCTCCCATTAATTGACTTTTGAA 1835
 601 HlsleubproasncysalaSerAlalaLeuaspheilleuysleuasphety 617
 1836 CATTTGCCAATTGTGCAAGTGTCTGAGACTCTAATTAAGTGCAGCTTTTA 1885
 617 rlyglyAlaMetAlaSerTrpGluYalaAlaGluAspThrGlyIYI 634
 1886 TGGGGGAGCTATGCTTCATGGGAAAAGCTGCAGAGACACAGCTGGAA 1935
 634 lelhsmetGluGluAlaProGluThrTrpIleProSerArgAlaValSer 650
 1936 TCCACATGGAGAGAGCCCGCAAAACCTACATTTCCAGCAGGCGCTTACT 1985
 651 LeuphepheaSnTrpLysGlnGluPheArgThrLeuGluValThrLeuAr 667
 1986 TTGTTCTTCAACTGGAAGCAGGAATTCAGACCTCTGGAGGTCAACCTCCG 2035
 667 gasPheSerLysLeuAsnLysGlnAspIleThrTrpLeuGlyLysIleP 684
 2036 GGATTTTCAGCAAGTTGAATAGCAAGATATCACATATCTGGGGAAATAT 2085
 684 heserSerAlaThrSerLeuArgLeuGlnIleLysArgCysAlaGluVal 700
 2086 TCAGCTTCCACACAGCCCTCAGCTGCACAAATTAAGAGATGTGCTGGTGT 2135
 701 AlaglySerLeuSerLeuValLeuSerThrCysLysAsnIleYerSerle 717
 2136 GCTGGAAGCCCTCAGTTTGTCTCAGCACCTGTAAAGAACATTTTCTCT 2185
 717 uMetValGluAlaSerProLeuThrIleGluAspGluArgHisIleThrS 734
 2186 CATGTGGAAGCCAGTCCCTCACCATAGAAAGATGAGAGCACATCACAT 2235
 734 erva1ThrAsnLeuLysThrLeuSerIleHisAspLeuGlnAsnGlnArg 750
 2236 CTGTAACAAACCTGAAAACCTGTAGTATTATGACTACAGAAATCAACGG 2285
 751 LeuProGlyLysLeuThrAspSerLeuGlyAsnLeuLysAsnLeuThrLy 767
 2286 CTGCGGGGTGTGACTGACAGCTGTGTAACCTTGAAGAACCTTACAAA 2335
 767 sleu1IleMetAspAsnIleLysMetAsnGluGluAspAlaIleLysLeuA 784
 2336 GCTCATATATGATTAACATTAAGATGAAGATGATATATAAACTAG 2385
 784 lagluGlyLeuLysAsnLeuLysLysMetCysLeuPheHisIleuThrHis 800
 2386 CTGAGGCCCTGAAAACCTGAAAGAGATGTGTTTATTTTCAATTGACCCAC 2435
 801 LeuSerAspIleGlyGluGlyMetAspTrpIleValLysSerLeuSerSe 817
 2436 TTGCTGTACATTTGGAGAGGATGATGATTAAGTCAAGTCTCTGTCAAG 2485
 817 rGluProCysAspLeuGluGluIleGlnLeuValSerCysCysLeuSerA 834
 2486 TGAACCCCTGTGACCTTGAGAAATTAATTAAGTCTCTGTCTGTGTCTG 2535
 834 laasna1ValLysIleLeuAlaGlnAsnLeuHisAsnLeuValLysLeu 850
 2536 CAATGACAGTGAATAATCTTACCTCAGATCTTCCACATTTGGTCAAACTG 2585
 851 Ser1LeuAspLeuSerGluAsnTrpLeuGluLysAspLysGlnGluAl 867
 2586 AGCATTTCTTGTATTCAGAAATTTACCTGGAAAAAGATGGAATGAAGC 2635
 867 aleuHisGluLeuLeuLysParGlyMetAsnValLeuGluGlnLeuThrAla 884
 2636 TCTTCATGAAGACTGATGACAGGATGAACGTGTAGAACACCTCACCCGAC 2685

884 eumelleProTrpGlyCysAspValGlnGlySerLeuSerSerLeu 900
 2686 TGATGCTGCCCTGGGCGCTGTGACCTGCCAAGCAGCTGAGACGCTTTG 2735
 901 LysHisLeuGluGluValProGlnLeuValLysLeuGlyLeuLysAsnTr 917
 2736 AAACATTTGGAGGAGGTGCCAACACTGCTCAGACTGTGGGTTGAAAACTG 2785
 917 ParGluThrAspThrGluIleArgIleLeuGlyAlaPhePheGlyLysA 934
 2786 GAGACTCACAGATACAGAGATTAGAAATTTAAGTGCAATTTTTCGAAAGA 2835
 934 snProLeuLysAsnPheGlnGlnLeuAsnLeuAlaGluAsnArgValSer 950
 2836 ACCCTTGAAAAACCTTCCAGCAGTTGAATTTGGCGGGAATCGTGTAGC 2885
 951 SerAspGlyTrpLeuAlaPheMetGlyValPheGluAsnLeuLysGlnLe 967
 2886 AGTGATGATGAGCTTGGCTTCATGTGGTGTATTGAGAAATCTTAAGCAAT 2935
 967 ValPhePheAspPheSerThrLysGluPheLeuProAspProAlaLeuV 984
 2936 AGCTTTTTCACCTTGTAGTAAAGAAATTCACATGATCCAGCAATAG 2985
 984 AlArgLysLeuSerGlnValLeuSerLysLeuThrPheLeuGlnGluAla 1000
 2986 TCAGAAACCTTAGCCAAAGTGTATCCAAAGTTACTTTTTCGCAAGAACT 3035
 1001 ArgLeuValGlyTrpGlnPheAspAspAspLeuSerValIleThrG1 1017
 3036 AGGCTTGTGGTGCCAAATTTGATGATGATGATGATGATGATGATGATG 3085
 1017 YAlaPheLysLeuValThrAla 1024
 3086 TGCTTTTAACTAGTACTGCT 3107
 seq_name: gb_pr:AY035391
 seq_documentation_block:
 LOCUS AY035391 3219 bp mRNA PRI 24-JUL-2001
 DEFINITION Homo sapiens ICE-protease activating factor mRNA, complete cds.
 ACCESSION AY035391
 VERSION AY035391.1 GI:14334214
 KEYWORDS
 SOURCE
 ORGANISM
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE
 1 (bases 1 to 3219)
 Fernandes-Alnemri,T. and Alnemri,E.S.
 Identification of Ipaf, a human caspase-1-activating protein
 related to Apaf-1
 J. Biol. Chem. 276 (30), 28309-28313 (2001)
 TITLE
 JOURNAL
 PUBMED 11390368
 MEDLINE 21359454
 JOURNAL
 AUTHORS
 REFERENCE
 2 (bases 1 to 3219)
 Poyet,J.-L., Srinivasula,S.M., Tnani,M., Raznaraj,M.,
 Alnemri,E.S.
 Direct Submission
 Submitted (16-MAY-2001) Microbiology and Immunology, Thomas
 Jefferson University, 233 S. 10th Street, Philadelphia, PA 19107,
 USA
 FEATURES
 source
 1..3219
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /map="p21-22"
 145..3219
 /note="IPAF: CED4/Apaf-1 family member; caspase-associated
 recruitment domain containing protein"
 /codon_start=1

```
/product-"ICE-protease activating factor"
/db_xref="GI:14334215"
/translation="MNEIKDNRALIQRMGTIVKQITDDLEFVNNVLRREVNITICE
KRVODARGLIIHMLKKGSESCNLIKEMNPILRODINGSLSPHOTSEGDIDLA
ODIKDYHPSEPLNFYPLGSDIDIDFNKLSTFEPPVIMRDOHHRRERQULNGLLDA
LQSPCITIEGSGKSKSTLLORIAMLWGSCKALTKRFYFELSLRAQGLRETTLCD
QLDIDPQIRKQTFPMAMLLKRNQVLELDGYNFKQNCPEILAKENHREKNMYI
VTTTECLRIHQGLTAEGMTEDSAQALIREVILKEIAELDIQISRLRLI
MPTPLFVITICAIOMSEFSPHSHQTLTFHTFYDLLKQKNHKKRGVAADEFISLH
CGDLAEGVFSHKDEFELQDVSSVNEEDVLTITGLCKVTLQRFKPKFKHROEYH
AGRRSLTSHSEPEYTKNGCYLOKHWSTSDTISYTSILRTYCGSSVATRAVYKH
LAAYVQGCILGLSIARPLMRQESLOSVAKNTBQELKAININSEVEGCIHLQEST
SKSALDSQEFALFQKSLVYNSGNIPOYLFDFEHLDPQNSALDFIKLDFEGAMASW
EKAEDDGEIHMEAPETYPISRAVSLFPMKQFRLTEVLRDPSLKNQIDQIYLIK
IFSSATSLRQIKRCAVASLSLVSTCKNIYSIAMEYKPLTJEDERHITSVNLKT
LSIHDONORLPGLTDSLGNLTKLIMDNIKMNEEDAKIAEGILKMKMLFHL
THSDIGEMDYIVKSLSPRCDFEYQIYSCLSAAVITLQNLNLVLYSLTDS
ENYLBKGNELHELDIDRNVLEQULALMLPWGDDVGSLSLKLHLEVPQVLKUL
KNMRLEDTETIRILGAFFGNPLKNFQDLNLNAGNVSSDGMIAFMGVFENKOLFDFD
STKEFLDPALVRLKLSQVLSKLPFLQEARLVGMQFDDDLSEVITGAKLVYA"
```

SE COUNT 934 a 696 c 748 g 841 t
IGIN

Alignment scores:
Quality: 5275.00 Length: 1024
Ratio: 5.156 Gaps: 0
Percent similarity: 99.902 Percent identity: 99.902

alignment_block:
US-09-697-089-2 x AY035391 ..

Align seg 1/1 to: AY035391 from: 1 to: 3219

```
1 Metasphelleielysaspasnserrglaaleuileglnargmetglyme 17
|||||
145 ATGAATTTTCATTAAGGACATAGCCGAGCCCTTATTCANAAGATGGAGAT 194
|||||
17 tthrval1lelysglnlethrasspaspheuphevaltrpasnvalleua 34
|||||
195 GACCTGTTTAAAGCAATCACAGATGACCTATTGTGATGGAATGTTCTGA 244
|||||
34 snargglugluvalasnlelleecysgsluylvalgluinalaspala 50
|||||
245 ATCCGGAAGAACTAAACATCATTTCTGCGAAGCTGAGCAGCATGCT 294
|||||
51 Alaargglyllellehismetlleuleuylsglyserglusercysas 67
|||||
295 GCTAGAGGATCATTCACATGATTTTGAANAAGGCTTCAGAGTCCCTGTA 344
|||||
67 nleupheuleuylserleuylsglutrpasntrprouleupheglnaspl 84
|||||
345 CCTCTTCTTAAATCCCTTAAAGAGTGAACATATCCTCATTTTCAGACT 394
|||||
84 euasnglyglinserleuphehisglntrsergluglyaspheuaspsp 100
|||||
395 TGAATGGAACAAGCTCTTTTTCATCAGACATCAGAAGGAGACTTGGACGAT 444
|||||
101 leuallaginaspleuleuylaspleuyltrhtrproserpheleuasph 117
|||||
445 TTGGCTCAGGATTTAAAGGACTTGTACCATACCCCATCTTTTCTGAACCT 494
|||||
117 euyrproleuglygluasplleaspllellepheasleuylserthr 134
|||||
495 TTATCCCTTGGTGAAGATATTCATATTTTAACTTGAAGAACACT 544
|||||
134 herthgluprovalleutrparglysasrglnhisishisargvalglu 150
|||||
545 TCACAGAACCTGTCTCTGTGAGGAAGAACCAACACCATCACCGCGTGAG 594
|||||
151 gluleuthrleuasnglyleuileuinalaleuulinserprocysille 167
|||||
```

```
595 CAGCTGACCCGTGAATGCGCTCTCGAGGCTTTCAGAGCCCTGCATCAT 644
167 egluglyluserglylysglyluserthrleuenglnargleualam 184
|||||
645 TGAAGGGAATCTGGCAAGCAAGTCCCTGCTGCGACCAATTTGCCA 694
|||||
184 etleuthrpllyserglylscylsalaaleuthrlyspheulyspheval 200
|||||
695 TCCTCGGGCTCCGGAAAGTCAAGGCTCTGACCAAGTCAATTCCTGTC 744
|||||
201 pheupheulargluserarglaaleuileuileuylphegluthrleucy 217
|||||
745 TTCTTCTCTCTGCTCAGCAGGGCCCAAGGGTGACCTTTTGAACCTCTG 794
|||||
217 saspglnleuleuasplleprogllythrlearglysglntrpmeteta 234
|||||
795 TCATCACTCCCTGGATATACCTCGCACATTCAGGAAGCAATTCATGG 844
|||||
234 lmetleuleuylsleuarglnargvalleupheuleuaspglytyr 250
|||||
845 CCATGCTGCTGAAGCTGCGGACAGAGGCTTCTTCTCTTGATGAGCTAC 894
|||||
251 asnglupheulysproglinasncysprogluileglnaleuileuylsgl 267
|||||
895 AATGAATTCAGCCCGCAGAACTGCGCAGAAATCGAAGCCCTGATTAAGA 944
|||||
267 uasnhisargpheulysasmetvallevalthrtrhtrthrnglucyl 284
|||||
945 AATCCACCGCTTCAGAAACATGTCATCTCACACCTACACTGAGTGGC 994
|||||
284 euarghis1learglnphegllyalaaleuthralaglvalglaspmet 300
|||||
995 TGAGGCACATACGCGCAGTTTGGTCCCTGACTGCTGAGGTGGGAGATATG 1044
|||||
301 thrgluaspserraglaaleuilearggluvalleuileuileuylsglue 317
|||||
1045 ACAGAAAGACAGCCCGCAGGCTCTCATCCGAGAAATGCTGATTAAGGAG 1094
|||||
317 ualaglulyleuleuleuileuileglnlysserrargcysleuargasnL 334
|||||
1095 TCCTGAAGCTTGTGCTCCAAATTCAGAAATCCAGAGTCTTGAGGAATC 1144
|||||
334 eumetlustrhrproleuphevalvallethr-cysala1leeglnmetgly 350
|||||
1145 TCATGAAGACCCCTCTCTTGTGTCATCATCTTGCAATCCAGATGGGT 1194
|||||
351 glusergluphehisserhisthrnglntrhrleuphehis1thrphe 367
|||||
1195 GAAAGTGAAGTCTCAGCTCACACACAAACAGCGTGTTCATACCTCTTA 1244
|||||
367 rasperleuleuileglnlysasnlyshislyshislysglyvalala1as 384
|||||
1245 TGATCTGTTGATACAGAAACAAACAAACATTAAGTGTGGGTGCA 1294
|||||
384 era1spheilleargserleuasphhis1cysglyaspheualaleuileugly 400
|||||
1295 GTGACTTCATTCGGAGCCCTGGACCACTGTGAGACCTGTGAGAGGT 1344
|||||
401 val1pneserhis1lyspheaspphegluleuileuinaspvaliser1valas 417
|||||
1345 GTGTTCTCCACAGATTTGATTTTGAACCTGACAGATGTGTCCAGGCTGA 1394
|||||
417 ngluaspyalleuleuthrthrthrglyleuleucyslystrprralaglna 434
|||||
1395 TGAGAGTGTCTGTCGACAACTGGGCTCCTCTGTAATATATACAGTCAAA 1444
|||||
434 rghelysprolystrlysrphepnehis1lysserphleglnlutythr 450
|||||
1445 GCTTCAAGCAAGATATAAATCTTTCAACAAGTCAATTCAGAGATACACA 1494
|||||
451 alaglyatrgargluserserleuuthrserhis1gluproglygluvala 467
|||||
1495 GCAGGACGAAGACTCAGCATTTATGAGTCTCATGAGCCGAGAGAGGT 1544
|||||
```

```

467  IThrLysGlyAsnGlyTyrLeuGlnLysMetValSerIleSerAspIleT 484
      |||||||
1545  GACCAAGGGGAATGTTACTTCGAGAAAATGTTCCATTTCGACATTA 1594
      |||||||
484  hrSerThrTyrSerSerLeuLeuArgTyrThrCysGlySerSerValGlu 500
      |||||||
1595  CATCCACTTAATAGCAGCTGCTCCGGTACACCTGCGGTCACTCTGTGGA 1644
      |||||||
501  AlaThrArgAlaValMetLysHisLeuAlaAlaValTyrGlnHisGlyC 517
      |||||||
1645  GCCACGAGGCGCTGTTATGAAACCTCGCACAGATGATCAACAGCGCTG 1694
      |||||||
517  sLeuLeuGlyLeuSerIleAlaLysArgProLeuTrpArgGlnGlnSerL 534
      |||||||
1695  CCTTCCTCGACTTTCATCGCCAGAGAGGCTCTCTGGAGACAGCAATTT 1744
      |||||||
534  euGlnSerValLysAsnThrThrGlnGlnGluLLeuLysAlaIleAsn 550
      |||||||
1745  TGCAAAGCTGAAAAACACCACTGAGCAAGAAATTCGAAAGCCATAAAC 1794
      |||||||
551  IleAsnSerPheValGluCysGlyIleHisLeuTyrGlnGlnSerThrSe 567
      |||||||
1795  ATCAATTCTTTTGAAGTGTGGCATCATTTATATCAAGAGATACATC 1844
      |||||||
567  rLysSerAlaLeuSerGlnGlnPheGluAlaPhePheGlnGlnLysSerL 584
      |||||||
1845  CAAATCAGCCCTGAGCCCAAGAAATTTGAAGCTTTCTTCAAGCTAAAGCT 1894
      |||||||
584  euTyrIleAsnSerGlyAsnIleProAspTyrLeuPheAspPheGln 600
      |||||||
1895  TATATATCAACTCAGGGAAATCCCGATCTACTTATTTGACTCTTTGAA 1944
      |||||||
601  HisLeuProAsnCysAlaSerAlaLeuAspPheIleLysLeuAspPheT 617
      |||||||
1945  CATTTGCCCAATTTGTCAAGTGCCCTGCACTTCATTTAACTGCACTTTTA 1994
      |||||||
617  rGlyGlyAlaMetAlaSerTrpGluLysAlaAlaGlnAspThrGlyLysI 634
      |||||||
1995  TGGGGAGACTTGCTGCTCATGGAAAAAGCTCCAGAGACACAGCTGCA 2044
      |||||||
634  LeHisMetGluGluAlaProGluThrTyrIleProSerArgAlaValSer 650
      |||||||
2045  TCCACATGAGAGAGGCCCCAGAAACCTACATTTCCAGACAGGCTGTATCT 2094
      |||||||
651  LeuPhePheAsnTrpLysGlnGluPheArgThrLeuGluValThrLeuAr 667
      |||||||
2095  TTGTTCTTCAACTGGAACAGGAATTCAGGACTCTGGAGGTCAACATCCG 2144
      |||||||
667  gAspPheSerLysLeuAsnLysGlnAspIleThrTyrLeuGlyLysIleP 684
      |||||||
2145  GGATTTTCAGCAAGTTGAATTAAGCAAGATATCATCTGGGGAAAAATAT 2194
      |||||||
684  heSerSerAlaThrSerLeuArgLeuGlnIleLysArgCysAlaGlyVal 700
      |||||||
2195  TCACCTCTGCACAAAGCTCAGGCTGCAAAATTAAGAGATGTGCTGTGTG 2244
      |||||||
701  AlaGlySerLeuSerLeuValLeuSerThrCysLysAsnIleTyrSerLe 717
      |||||||
2245  GCTGGAAACCTCAGTTGGTCTCTCAGCACCGTGAAGAACATTTATTCCT 2294
      |||||||
717  uMetValGluAlaSerProLeuThrIleGluAspGluArgHisIleThrS 734
      |||||||
2295  CATGCTGGAACCAAGTCCCTCACCATTAAGAGATGAGAGGACATCATAT 2344
      |||||||
734  euValThrAsnLeuLysThrLeuSerIleHisAspLeuGlnAsnGlnArg 750
      |||||||
2345  CTGTAACAACCACTGAAAACTTGAAGTATTCATGACCTACAGAACCAACG 2394
      |||||||
751  LeuProGlyGlyLeuThrAspSerLeuGlnLysLeuLysAsnLeuThrLys 767
      |||||||
2395  CTGCGCGGTGTGTGACTGACAGCTTGGTGAACCTTGAAGAACCTTACAA 2444
      |||||||

```

```

767  sLeuIleMetAspAsnIleLysMetAsnGluGluAspAlaIleLysLeuA 784
      |||||||
2445  GCTCATATATGATTAACATAAAGATGAATGAAGAAGATGCTATAAATACTAG 2494
      |||||||
784  IaGluGlyLeuLysAsnLeuLysLysMetCysLeuPheHisLeuThrHis 800
      |||||||
2495  CTGAAGGCCCTGAAAAACCTGAAGAAGATGTGTTATTTCAATTGACCCAC 2544
      |||||||
801  LeuSerAspIleGlyGluGlyMetAspTyrIleValLysSerLeuSerSe 817
      |||||||
2545  TTGTCTGACATTTGAGAGGGAATGATTAATCAATCAAGTCTGTCAAG 2594
      |||||||
817  rGluProCysAspLeuGlnGluIleGlnLeuValSerCysCysLeuSerA 834
      |||||||
2595  TGAACCCGTGACCTTGAGAAATTCATTTGTTCTCTCGCTGCTGTCTG 2644
      |||||||
834  IaAsnAlaValLysIleLeuAlaGlnAsnLeuHisAsnLeuValLysLeu 850
      |||||||
2645  CAAATGCGAGTGAATAATCTAGCTCAGAAATCTTCACAATTTGTCAAACTG 2694
      |||||||
851  SerIleLeuAspLeuSerGluAsnTyrLeuGlnLysAspGlyAsnGlnAl 867
      |||||||
2695  AGCATTTCTTGAATTTATCAGAAATTTACCTGGAAAAAGATGCAATGAAGC 2744
      |||||||
867  aLeuHisGluLeuLeuIleAspArgMetAsnValLeuGlnGlnLeuThrAl 884
      |||||||
2745  TCTTCATGAACTGATTCAGACAGATGAACGTGCTGAACAGCTCACCCGAC 2794
      |||||||
884  euMetLeuProTrpGlyCysAspValGlnGlySerLeuSerSerLeuLeu 900
      |||||||
2795  TGATGCTGGCCCTGGGGCTGTGACGTGCAAGGACACCTGAGAGGCTGTG 2844
      |||||||
901  LysHisLeuGlnGluValProGlnLeuValLysLeuGlyLeuLysAsnTr 917
      |||||||
2845  AAACATTTGGAGGAGGTCCCAACACTGCTCAAGCTGTGGTTGAAAACTG 2894
      |||||||
917  ParGluThrAspThrGluIleArgIleLeuGlyAlaPhePheGlyLysA 934
      |||||||
2895  GAGACTCAGACATTCAGAGATTAAGATTTTAGTGCAATTTTGTGAAAGA 2944
      |||||||
934  snProLeuLysAsnPheGlnGlnLeuAsnLeuAlaGlyAsnArgValSer 950
      |||||||
2945  ACCCTCTGAAAAAATCTCCAGCAGTGAATTTGGCGGAAAAATCGGTGAGC 2994
      |||||||
951  SerAspGlyTrpLeuAlaPheMetGlyValPheGluAsnLeuLysGlnLe 967
      |||||||
2995  AGTATGGAATGGCTTGCCTTCATGGGTATTTGAGAATCTTTAAGCAATT 3044
      |||||||
967  uValPhePheAspPheSerThrLysGluPheLeuProAspProAlaLeuV 984
      |||||||
3045  AGTGTTTTTTCACCTTTAGTACTAAAGAATTTCTACCTGATCCAGCATTAG 3094
      |||||||
984  aLArgLysLeuSerGlnValLeuSerLysLeuThrPheLeuGlnGluAla 1000
      |||||||
3095  TCAGAAAAAATTAAGCAAGTGTATTCACCAAGTTACTTTCTGCAAGAGCT 3144
      |||||||
1001  ArgLeuValGlyTrpGlnPheAspAspAspAspLeuSerValIleThrG 1017
      |||||||
3145  AGGCTTTTGGGTGGCAATTTGATGATGATGATCTCAGTGTATTTACAGC 3194
      |||||||
1017  yAlaPheLysLeuValThrAla 1024
      |||||||
3195  TGCCTTTAACTAGTAATCTGCT 3216
      |||||||

```

seq_name: gb_pr:AF376061

seq_documentation_block: 3581 bp

LOCUS AF376061 mRNA 15-MAY-2001

DEFINITION Homo sapiens caspase recruitment domain protein 12 mRNA, complete

CDs: AF376061
 ACCESSION AF376061
 VERSION AF376061.1 GI:14040074
 KEYWORDS

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 3581)
AUTHORS Gingras, M., Qiu, J., and Margolin, J.F.
TITLE Differential expression of the caspase recruitment domain protein
12 (CARD12) during monocytic differentiation
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 3581)
Gingras, M., Qiu, J., and Margolin, J.F.
TITLE Direct Submission
JOURNAL Submitted (03-MAY-2001) Pediatric/Texas Children's Cancer Center,
Baylor College of Medicine, 6621 Fannin St., MC3-3320, Houston, TX
77030, USA
FEATURES
SOURCE location/qualifiers
1..3581 /organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="3p21-22"
/cell_line="0937"
/cell_type="peripheral blood-derived CD14 mature monocyte"
490..3564 /note="CARD12"
/product="caspase recruitment domain protein 12"
/protein_id="AAK53443.1"
/db_xref="GI:14040075"
/translation="MNFIKDSRALIORMKNTVIKQITDDLEFVNVNLRNEVNIICE
KVEDDAARGIIHMLIKGSESCNLFSLKEMWPLFDLNGOSLFQTSQGDLDLA
QDLKDLVHTSFNFYPLGEDIIDIFLKSFTFPLVMRKQDHHRYEOLNLGLDA
LOSFCILEGSGKSKYLORIAMLSGSKKALTKRFVFEFLSHAGGLFETLDD
OLDIDPGTIRKOTFMAMILKRLORVFLFDGYNFKNQONDEIHALIKENHRENMVI
VTPTTECRHTRGOCALFAYGSDTSACALILEVILKELAESLLIQIKSRRLNL
MKTPLEVITICAIOMGESERHSHTOTTLFTFTDGLIÖKNKHKGVAASDFISLDH
CGDLLEGVFSHKDFELQDVSYNEDALLTGTGLCKVTAGREFPKYKFEKFSQEXT
AGRLSILTSISHEPEVTNGVYLOKNVSIISDITSTYSLRTYCGSSVEATRAVMH
LAAVYOHGCLGLSILAKRPLROESLOSVMKTTQETILKALINISFVCGIHLTQEST
SKALSOEPEAFQOGKSLYINGNIPLYLPEFPHLNCASALDFIKLDFEGGMAAM
EKAEADGTGIMEAPEPTIIPSRVSLFPMNKQEPFLREVLTRFSKLNODIRYLK
IFSSATSLRLQIKRCAGVAGSLVLTCKNTIYSLMEASFTITEDERHITVTNLT
LSTHDLONRLPGELTDSLGNLKNLTKLMDNIMNEEDAKIKLAEGLNKKWLFH
TLHSDIGENMDYIVKSLSEPCDLEIQLVSCSANAAYKILAEGLNKKWLFH
ENYLEKGDNEILHELDIMNVLEQLTALMLPGCDVGSLSLKHLEEVYQLVKGL
KNMRLDTEIRILGAFRGNPLKNFOOLNLAGNVSDGMLAFGVPEENKOLVFPDF
STKEPLDPALVRKLSVLKTLFLQEARLVGMQFDDDLISVITGARKLVTYA"
misc_feature 502..741
/note="Region: caspase recruitment domain"
CDS
USE COUNT 1033 a 781 c 843 g 924 t
ALIGNMENT-scores:
Quality: 5271.00 Length: 1024
Ratio: 5.152 Gaps: 0
Percent Similarity: 99.902 Percent Identity: 99.805
alignment block:
US-09-697-089-2 x AF376061 ..
Align seg 1/1 to: AF376061 from: 1 to: 3581
1 MetAsnPheIleLysAspAsnSerATgAlaLeuIleGlnArgMetGlyLeu 17
|||||
490 ATGAAATTTTATAAGGACAATAGCCGACCTTATTTCAGAGATGGCAAT 539
17 tThrValIleLysGlnIleThrAspAsnLeuPheValITrPAsnValLeuA 34
|||||
540 GACTGTATATAAGCAATCACAAGATGACCTATTGTGATGAAATGTTCTGA 589
34 snArgGluGluValAsnIleIleCysGlyLysValGluGlnAspAla 50
|||||

590 ATCGCGAAGAGTAACATCATTTGCTGCGAGAGAGGTGAGCAGATGCT 639
51 AlArGGLyIleIleHisMetIleLeuLysGlySerGluSerCysAs 67
|||||
640 GCTAGAGGATCATTCACATGATTTTGAAAAGGGGTGAGATCCTCTGA 689
67 nLeuPheLeuLysSerLeuLysGluITrPAsnITrProLeuPheGlnAspL 84
|||||
690 CCTCTTTCTTAATCCCTTAAGAGAGTGAGACATATCCTCATTTTCA 739
84 euAsnGlyGlnSerLeuPheHisGlnITrSerGlyLysAspLeuAsp 100
|||||
740 TGAATGACAAAGTCTTTTTCATCAGACATCAGAGAGACTTGGACAT 789
101 LeuAlaGlnAspLeuLysAspLeuITrHisITrProSerPheLeuAsnPh 117
|||||
790 TTGGCTCAGGATTTAAAGACTGTGACATACCCATCTTTCTGAACCT 839
117 eITrProLeuGlyGluAspIleAspIleIlePheAsnLeuLysSerITrP 134
|||||
840 TTATCCCTTGTGTAAGATATGACATTAATTTTAACTTGAAGACACCT 889
134 heITrGluITrProValLeuITrPArgLysAspGlnHisITrHisITrValGlu 150
|||||
890 TCACAGAACCACTCCTGTGAGAGAACCAACACCATCACCCTGTGAG 939
151 GlnLeuITrLeuAsnGlyLeuLeuGlnAlaLeuGlnSerProCysIleI 167
|||||
940 CAGCTGACCTCAATGCTCCTCGACAGCTCTTCAGAGCCCTGCATCAT 989
167 eGluGlyGlnSerGlyLysGlyLysSerITrLeuLeuGlnArgIleAla 184
|||||
990 TCAAGAGGAAATCGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1039
184 eITrLeuITrPArgLysGlyLysGlyLysAlaLeuITrLysPheLysPheVal 200
|||||
1040 TCCTCTGGGGCTCGGAAAGTCAAGGCTGTGCAAGGTTCAAAATTCGTC 1089
201 PhePheLeuITrLeuSerITrArgAlaGlnGlyLysLeuPheGluITrLeu 217
|||||
1090 TTCTTCTCCTCGCTCAGCAGGAGGCTGAGCTTTTGAACCTCTG 1139
217 sAspGlnLeuLeuAspIleITrProGlyITrITrLeuArgLysGlnITrPheMet 234
|||||
1140 TCATCAACTCTCGATATACCTGACCAATCAGGAAGCAGACATTCATG 1189
234 lAmelLeuLeuLysLeuITrArgGlnITrValLeuPheLeuLeuAspGlyITr 250
|||||
1190 CCATGCTGCTGAAGCTGCGGAGAGGTTCTTCTCTTGATGGCTAC 1239
251 AsnGluPheLysProGlnAsnCysProGluITrGlnAlaLeuIleLysG 267
|||||
1240 AATGAATTTCAAGCCCGCAGAACTGCCAGAAATCGAAGCCCTATTAAGA 1289
267 uAsnHisITrPheLysAsnMetValIleValITrITrITrITrGluCysL 284
|||||
1290 AAACCAACCGCTTCAAGAACATGTCATGCTCACACATCACATGAGTGC 1339
284 euITrHisITrLeuITrGlnPheGlyAlaLeuITrITrAlaGluValGlyAsp 300
|||||
1340 TGAAGCACATACGAGGATTTGGTGGCTGTGCTGTGAGGTGGGATTAAG 1389
301 TrnGluAspSerITrAlaGlnAlaLeuIleArgGluValLeuIleLysG 317
|||||
1390 ACAGAAAGAGAGCGCCAGGCTCTCATCCAGAGAGTGCATCAAGAGAGCT 1439
317 uAlaGluGlyLeuLeuGlnIleGlnLysSerITrArgCysLeuITrArgAsnL 334
|||||
1440 TCGTGAAGCTGTGTGCTCAATTCAGAAATCCAGGTGCTTGAAGAAATC 1489
334 euMetLysITrITrProLeuPheValIleITrITrCysAlaIleGlnMetGly 350
|||||
1490 TCATGAAGACCCCTCTTTTGTGTATCATCATTTGTGCAATTCAGATGGGT 1539

351 GluSerGluPheHisSerHisThrGlnThrThrLeuPheHisThrPheTy 367
1540 GAAAGTAGTCCACTCTCACACAAACAACGCTGTTCCATACCTTCTA 1589
367 rAsPLeuLeuIleGlnLysAsnLysHisLysHisLysGlyValAlaIAs 384
1590 TGATCTGTGATACAGAAAAACAACAACATAAAGGTGGCTGCACAA 1539
384 eArSPheIleArGSerLeuAspHisCysGlyAspLeuAlaLeuGly 400
1640 GTACTCTCATTCGAGCGCTGACACACTGTGAGACCTACCTCGAGAGGT 1689
401 ValPheSerHisLysPheAspPheGluLeuGlnAspValSerSerValAs 417
1690 GTGTCTCCCAACAAGTTTGATTTGCAACTCGAGATGTGTCACGCTGAA 1739
417 nGluAspValLeuLeuThrThrGlyLeuLeuCysLysTyThrAlaGln 434
1740 TGAAGATGCCCTGCTGACACTGGGCTCCTGTGAATATACAGCTCAAA 1789
434 rPheLysProlTyThrLysPhePheHisLysSerPheGlnLysTyThr 450
1790 GGTTCACAGCCAAAGTAAATCTTTCACAAGCATTCACAGAGTACACA 1839
451 AlaGlyArgArgLeuSerSerLeuLeuThrSerHisGluProGluGlu 467
1840 GCAGGACGAAGACCTACAGCTATTGACCTGTATAGCCACAGAGGAGGT 1889
467 LTrhLysGlyAsnGlyTyThrLeuGlnLysMetValSerLysSerAspIleT 484
1890 GACCAAGGGAATGGTTACTTGCAGAAATGGTTCCATTTGCGACACTTA 1939
484 hSerThrTyThrSerSerLeuLeuArgTyThrCysGlySerSerValGlu 500
1940 CATCCACTTATACAGCGCTGCTCGGTACACCTGTGGTCATCTGTGAAA 1989
501 AlaThrArgAlaValMetLysHisLysLeuAlaValTyGlnHisGlyC 517
1990 GCCACCGGGCTGTATGAAGCACCTCGACAGCTGTATCAACACGGCTG 2039
517 sLeuLeuGlyLeuSerIleAlaLysArgProLeuTrpArgGlnLysSerL 534
2040 CCTTCTCGACTTTCATCGCCAAAGAGCGCTCTGTGAGACAGGAATCTT 2089
534 euGlnSerValLysAsnThrThrGlnGlnGluIleLeuLysAlaIleAsn 550
2090 TGCAAAGTGTGAAAAACACACACGTGACAGCAAAATCTGAAAGCCATPAAAC 2139
551 IleAsnSerPheValGluCysGlyIleHisLeuTyrglnGlnLysSerPhe 567
2140 ATCAATTCCTTTAGAGTGTGCATCCATTTTATCAAGAGAGTACATC 2189
567 rLysSerAlaLeuSerGlnGluPheGluAlaPhePheGlnGlyLysSerL 584
2190 CAATATGAGCCCTGAGCCCAAGAATTGAAGCTTCTTCAAGGTAAAAAGCT 2239
584 euTyrrIleAsnSerGlyAsnIleProAspTyrrLeuPheAspPheGlu 600
2240 TATATATCAACTCAGGGAACATCCCGATTAATTGACTTCTTTGAA 2289
601 HisLeuProAsnCysAlaSerAlaLeuAspPheIleLysLeuAspPheTy 617
2290 CATTTGCCCAATTTGTCAAGTGTGCCCTGTGACTTCAITTAAGTGAATTTTA 2339
617 rGlyGlyAlaMetAlaSerTrpGluLysAlaAlaGluAspThrGlyLys 634
2340 TGGGGGAGCTATGCTCATGGGAAAAAGCTGCAGAAACACAGGTGCA 2389
634 LeHisMetGluGluAlaProGluThrTyrrIleProSerArgAlaValSer 650
2390 TCACATGGAAGAGGCCCAAGAAACCTACATTCGCCAGAGGCTGTATCT 2439

651 LeuPheAsnTrpLysGlnGluPheArgThrLeuGluValThrLeuArg 667
2440 TTGTCTTCACTGGAAGCAGGAATTCAGAGCTGTGAGGTCACTCCG 2489
667 gaSPheSerLysLeuAsnLysGlnAspIleThrTyrrLeuGlyLysIleP 684
2490 GCATTTCAAGAGTTGAATAGCAAGATATCAGATATCTGTGGGAAAAATAT 2539
684 hSerSerAlaThrSerLeuArgLeuGlnIleLysArgCysAlaGlyVal 700
2540 TCAGCTCTGCCAACCCCTCAGCTCCTCAATTAAGAGATGTGCTGTGTG 2589
701 AlaGlySerLeuSerLeuValLeuSerThrCysLysAsnIleTySerLe 717
2590 GCTGGAAGCTCAGTTTGTCTCAGACACCTGTAAACAATTTATCTCTG 2639
717 uMetValGluAlaSerProLeuThrIleGluAspGluArgHisIleThrS 734
2640 CATGGTGAAGCCAGTCCCTCACCATTAGAGATGAGAGGCACATCAAT 2689
734 eValIleHisAsnLeuLysThrLeuSerIleHisAspLeuGlnAsnGlnArg 750
2690 CTGTAAACAAACCTGAAGAACCTTGATATTCATGACCTACAGAAATCAACG 2739
751 LeuProGlyGlyLeuThrAspSerLeuGlyAsnLeuLysAsnLeuThrLy 767
2740 CTGCGGGGTGTCTGACTGACACCTTGGGTAACTTGAAGACCTTCAAA 2789
767 sLeuIleMetAspAsnIleLysMetAsnGlnGluAspAlaIleLysLeu 784
2790 GCTCATATAGTATACATATAAGATGAATGAAGAATGCATATAAACTAG 2839
784 laGluGlyLeuLysAsnLeuLysLysMetCysLeuPheHisLeuThrHis 800
2840 CTGAAGCCCTGAAAAACCTGAAGAAGATGTGTTATTTCAATTGACCCAC 2889
801 LeuSerAspIleGlyGlnGlyMetAspTyrrIleValLysSerLeuSer 817
2890 TTGTCTGACATTGGAGAGGAATGGATTTACATAGTCAAGTCTTGTCAAG 2939
817 rGluProCysAspLeuGlnGluIleGlnLeuValSerCysLysLeuSer 834
2940 TGAACCTGTGACCTTGAAGAAATTCATATAGTCTGCTGTCTGTG 2989
834 laAsnAlaValLysIleLeuAlaGlnAsnLeuHisAsnLeuValLysLeu 850
2990 CAATGTCAGTGAATAATCTTACAGATCTTCACAATTTGGTCAAACTG 3039
851 SerIleLeuAspLeuSerGluAsnTyrrLeuGluLysAspGlyLysGln 867
3040 AGCATTTCTGATTTATCAGAAAATTTACTGTGAAAAAGATGGAATAGAC 3089
867 aLeuHisGluLeuIleAspArgMetAsnValLeuGlnGlnLeuThrAla 884
3090 TCTTCATGAACGTATGACAGATGAACGTCAGAAACAGCTCACCGCAC 3139
884 euMetLeuProTrpGlyCysAspValGlnGlySerLeuSerSerLeu 900
3140 TGATGTGCTGCTGCTGCTGTGAGCTGCAAGGCGCTGTGACACCTGTG 3189
901 LysHisLeuGlnGluValProGlnLeuValLysLeuGlyLeuLysAsnTr 917
3190 AAACATTTGGAGAGGTCCACACACTGCTCAAGCTTGGGTGAAAAACTG 3239
917 rArgLeuThrAspThrGluIleArgIleLeuGlyAlaPhePheGlyLys 934
3240 GAGACTCACAGATACAGAGATTTGAATTTAGGTGATTTTGGAAAGA 3289
934 snProLeuLysAsnPheGlnGlnLeuAsnLeuAlaGlyAsnArgValSer 950
3290 ACCCTCTGAAAACTTCCAGCACTGTAATTTGGCGGGAATCGTGTGAGC 3339
951 SerAspGlyTrpLeuAlaPheMetGlyValPheGluAsnLeuLysGln 967


```

|||||
3340 AGTGTGATGCGCTTGCCTTCATGGGCTATTGTGAGAACTTTAAGCAAT 3389
967 uValPhepHeaspPheSerThrLySGluPheLeuProaspProAlaLeuV 984
3390 AGTGTTTTGTACTTACTACTAAAGAAATTTCTACTGATCCAGCATTTAG 3439
984 aAlargLysLeuSerGlnValLeuSerLysLeuThrPheLeuGlnGluAla 1000
3440 TCAGAAACTTAGCCAGAGTGTATCCAGTTTCTTCGACAGAGCT 3489
1001 ArgLeuValGlyTrpGlnPheaspPaspPaspLeuSerValIleThrG1 1017
3490 AGCGTTGTTGGTGGCAATTGATGATGATCATCAGTGTATTATACAG 3539
1017 yAlaPheLysLeuValThrAla 1024
3540 TCGCTTTAACTAGTACTGCT 3561
seq_name: gb_pr:AY027787

```

documentation_block:

```

JUS AY027787 3396 bp mRNA PRI 20-JUL-2001
DEFINITION Homo sapiens CLANA (CLAN1) mRNA, complete cds.
ACCESSION AY027787
VERSION AY027787.1 GI:14324112
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 3396)
Damiano,J.S., Stehlik,C., Pio,F., Godzik,A. and Reed,J.C.
JOURNAL
Genomics. 75 (1-3), 77-83 (2001)
MEDLINE
21365712
PUBMED
11472070

```

```

REFERENCE
2 (bases 1 to 3396)
Stehlik,C., Damiano,J.S., Pio,F., Godzik,A. and Reed,J.C.
AUTHORS
Submitted Submission
TITLE
Submitted (21-FEB-2001) Program on Apoptosis and Cell Death
JOURNAL
Research, The Burnham Institute, 10901 North Torrey Pines Road, La
Jolla, CA 92037, USA

```

FEATURES

```

source
1. .3396
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2p22-p21"
/lissue_type="lung"
1. .3396
/gene="CLAN1"
/gene="CLAN1"
277. .3351
/gene="CLAN1"
/codon_start=-1
/product="CLANA"
/protein_id="AAK14776.1"
/db_xref="GI:14324113"
/translatation="MNEIKDNRSLIORMGMTVYIKOITDLEFVNVNLRREVITICE
KYBODARGIHHILTKKSSPCNLFLSKENMYPIFLDINGSLTHQTSSEGLDILA
ODKLDIHTPSTINFTPLGEDIDITINLSTFTPEPLMKRDQHHHVEQITLGLLOA
LOSPLIIEBESGKSTLQRIAMLGSGKALTKFKFVFLRSRAOGLEETLCD
QULIDPGTRKQTFMAMLLKROVLFLDGVNEFPQNPETELIKEHREKNNVY
VTTTECLRHIOFGALTAEGDMTEDSAOLIREVILKELAGLILLOKSKCLANL
MKTPLEVTICATOWGSEPHSHOTTFTEFTVDLILOKKHKKHKGVAASDFRSDH
RGDLAEVPSHKEDELDVSVNEDVLTLLTKYTRQKRPKYKFKHKSQET
AGRRISLLTSHPEEVTGNGILOKAVSISDITSTYSLRTYCGSSVEATPAVKH
LAAYVHGCLLSIEIAKRPIMROESLOSVKNTQETLKAININSVEGCIHLYOEST
SKSALSOEFAFQGSILYINSINIDYLFDEFENLPNCASALDFIKLIDFYGAMSM
EKAAEDTGGIHEADEPETYIPRAVSLFENMKOEFTLEVTLEDFKLNKODIRYLGK
IFSSATSLRLOIRKRGAVAGSLVSTCKNYSILVAEASPLTIEDERHTSTNLT
LSTHDLOQRLOPGLTDSIGNLKNLTKLIMDNITKNMEEDAIIKABEGLNKLKCLFHL
THLSDIGEGMDYIVKLSLSPCDLIEIQLVSCCLSANAVKIIAQNHLNIVKLSILDLS

```

gene
CDS

```

BASE COUNT 992 a 737 c 793 g 874 t
ORIGIN
ENYLEKDGNEALHELIDRNNVLEQLTALMLPMWCGDVQSSLSLKLHEVPOLVKLGL
KNMRLLDTEIRILGAFNRGNPKLNFKPLNKVAGLNSDGLAFMVFENIKQLVFEDF
STKEFLPDPALVKKILSOVLSTKFTFLQEARLVGMQFDDDLIVITGAFKLVTA"

```

alignment_scores:
Quality: 5262.00 Length: 1024
Ratio: 5.149 Gaps: 0
Percent Similarity: 99.805 Percent Identity: 99.707

alignment_block:
US-09-697-089-2 x AY027787 ..

Align seg 1/1 to: AY027787 from: 1 to: 3396

```

1 MetanpHei1eLysAspAsnSerArgAlaLeuIleGlnAtgMetGlyMe 17
|||||
277 ATGAATTCATAAAGACAAATAGCCGACCCCTTATTTCAAAGATGGGAT 326
17 tThrValIleLysGlnIleThrAspAspLeuPheValIrrPasnValLeuA 34
|||||
327 GACTTTTATTAAGCAATACACGATGACCTATTGTATGGAATGTTCTGA 376
34 snATgGluGluValasnIleIleCysCysGluLysValGluGlnAspAla 50
|||||
377 ATCGGAGAAGAGTAAACATCATTTGCTCGAGAAGGTGAGAGAGATGCT 426
51 AlaATgGlyIleIleIleHisMetIleLeuLysGlySerGlySerCysAs 67
427 GCTAGAGGATCATTCACATGATTTTGAAGAGGTTGAGAGCTCTGTAA 476
67 nLeuPheLeuLysSerLeuLysGluTrpPasnTyrrProLeuPheGlnAspL 84
|||||
477 CCTCTTTTAAATCCCTTAAAGAGTGAACATTCCTTATTTACAGACT 526
84 euAsnGlyGlnSerLeuPheHisGlnThrSerGluGlyAspLeuAspAsp 100
527 TGAATGACAAAGTCTTTTTCATCAGACATCAGAAGGAGACTTGGAGAT 576
101 LeuAlaGlnAspLeuLysAspLeuTyrrHisThrProSerPheLeuAsnph 117
577 TTGGCTCAGATTAAAGACCTTGACATACCCCATCTTTTTCGAACCT 626
117 eTyrrProLeuGlyLysAspIleAspIleIlePheAsnLeuLysSerThrP 134
|||||
627 TTATCCCTTGCTGAGATATTGACATTATTTTACTTGAAGACACT 676
134 herhGluProValLeuThrParGlyAspGlnHisHisArgValGlu 150
|||||
677 TCACAGAACTATCCTCTGGAGGAAGCAACACCATCACCAGGTGAG 726
151 GlnLeuThrLeuAsnGlyLeuLeuGlnAlaLeuGlnSerProCysIleI 167
727 CAGCTGACCTGAAATGGCTCTCTGAGGCTCTTCAGACCCCTGCATCAT 776
167 eGluGlyGlnSerGlyLysGlyLysSerThrLeuLeuGlnArgIleAlaM 184
|||||
777 TGAAGGGAATCTGGCAAGGCAAGCTCCACTCTGCTGAGCGCATTTGCCA 826
184 etLeuTrpGlySerGlyLysCysLysAlaLeuThrLysPheLysPheVal 200
827 TGCCTGGGGCTCCGGAAGAGCAAGGCTCGACCAACTTCAAAATTCGTC 876
201 PhePheLeuArgLeuSerArgAlaGlnGlyLysLeuPheGlyThrLeuCy 217
877 TTCTTCCCTCCGCTCAGCAGGAGGCGGAGGCTTTTGAAGACCTCTG 926
217 sasPheLeuLeuAspIleProGlyThrIleArgGlyGlnThrPheMetA 234
927 TGATCAACTCTCGATTAACCTGGGACATCAGAGGAGGACGACATTCATG 976

```



```

234 IamLeuLeuLysLeuArgGlnArgValLeuPheLeuLeuAspGlyTyr 250
|||||
977 CCATGCTCTGTAAGCTGGCGAGAGGGTTCTTTCTTCTTGATGGCTAC 1026
|||||
251 AsnGlnPheLysProGlnAsnGlyProGluIleGluAlaLeuIleLysGln 267
|||||
1027 AATGAATTCAAGCCCGACAGAACTGCCAGAAATCGAAGCCCTGATAAAGA 1076
|||||
267 uAsnHisArgPheLysAsnMetValIleValIleThrThrThrGlnLysL 284
|||||
1077 AAACCAACCGCTTCAAGAACATGTCATGTCACCACTACCACTAGTGC 1126
|||||
284 euArgHisIleArgGlnPheGlyAlaLeuThrAlaGluValGlyAspMet 300
|||||
1127 TGAGGCAATACGGCAGATTGGTGCCTGACTGCTGAGGTGGGGAATAG 1176
|||||
301 ThrGluAspSerAlaGlnAlaLeuIleArgGluValLeuIleLysGluLe 317
|||||
1177 ACAGAAACACAGCCCGCAGGCTCTCATCCGAGAACTGTCATCAAGGAGCT 1226
|||||
317 uAlaGlnGlyLeuLeuLeuGlnIleGlnLysSerArgCysLeuArgAsnL 334
|||||
1227 TGTCTGAAGGCTTGTGCTCCAAATTCAGAAATCCAGGTGCTTGAGGATC 1276
|||||
334 euMetLysThrProLeuPheValValIleThrCysAlaIleGlnMetGly 350
|||||
1277 TCATGAAGACCCCTCTTGTGTGTCATCACTTGTGCAATCCAAATGGGT 1326
|||||
351 GluSerGlnPheHisSerHisThrGlnThrThrLeuPheHisThrPheTy 367
|||||
1327 GAAAGTAGTCCACTCTCACACAAACAAACGCTGTTCATACCTTCTTA 1376
|||||
367 rAspLeuLeuIleGlnLysAsnLysHisLysHisLysGlyValAlaIleAs 384
|||||
1377 TGATCTGTCATACGAAACAAACAAACAAACATAAAGGTGTGCTGCA 1426
|||||
384 erAspPheIleArgSerLeuAspHisCysGlyAspLeuAlaLeuGluGly 400
|||||
1427 GTGACTTCATTCGGAGCCTGGACACCGTGAAGACTAGCTGTGAGAGGT 1476
|||||
401 ValPheSerHisLysPheAspPheGluLeuGlnAspValSerSerValAs 417
|||||
1477 GTGTTCCTCCACAAGTTGATTCGAACCTCAGAGATGTGTCCACCGTGA 1526
|||||
417 nGluAspValLeuLeuThrThrGlyLeuLeuLysLysThrThrIleGln 434
|||||
1527 TGAGGATGTCCTGCTGAACCTGGGCTCTCTGTAATATATACAGCTAAA 1576
|||||
434 rGpPheLysProLysThrLysPhePheHisLysSerPheGlnGluTyrThr 450
|||||
1577 GGTTCAGACCCAAAGTATAAATTTCTTCACAAGTCATTCAGAGATACACA 1626
|||||
451 AlAGlyArgArgLeuSerSerLeuLeuThrSerHisGluProGluGluVal 467
|||||
1627 GCAGAGACGAAGACTCACAGTTTATGACGTCATGACGACAGAGGAGGT 1676
|||||
467 lThrLysGlyAsnGlyTyrLeuGlnLysMetValSerIleSerAspIleT 484
|||||
1677 GACCAAGGGAGATGGTTACTTGCAAAATGAGTTTCCATTTGGACATTA 1726
|||||
484 hrSerThrTyrSerSerLeuLeuArgTyrThrCysGlySerSerValGlu 500
|||||
1727 CATCCACTTATAGCAGCCTGCTCCGGTACACCTGTGGTCATCTGTGAA 1776
|||||
501 AlaThrArgAlaValMetLysHisLeuAlaAlaValTyrGlnHisGlyC 517
|||||
1777 GCCACAGGGCTGTATGAAGCACCTGCAGCACTGATCAACACGGCTG 1826
|||||
517 sleuLeuGlyLeuSerIleAlaLysArgProLeuTyrArgGlnGluSerL 534
|||||
1827 CTTTCTGGACTTTCATCCGCAAGAGGCTCTCTGTGAGACAGGAATCTT 1876
|||||
534 euGlnSerValLysAsnThrThrGlnGlnGluIleLeuLysAlaIleAsn 550
|||||
|||||
1877 TGCAAACTGTGAAAAACACCACCTAGCAGAAATAATTCGAAAGCATTAAC 1926
|||||
551 lIleAsnSerPheValGluCysGlyIleHisLeuTyrGlnLysThrSe 567
|||||
1927 ATCAATTCCTTTGTAGAGTGTGGCATCTTATATCAAGAGATATATC 1976
|||||
567 rLysSerAlaLeuSerGlnGluPheGluAlaPhePheGlnGlyLysSerL 584
|||||
1977 CAAATCAGCCCTGAGCCACAAATTGTAAGCTTCTTTCAGAGTAAAGCT 2026
|||||
584 euTyrIleAsnSerGlyAsnIleProAspTyrLeuPheAspPhePheGlu 600
|||||
2027 TATATATCAACTCAGAGACATCCCGATTAATTAATTTGACTTCTTGAA 2076
|||||
601 HisLeuProAsnGlyAlaSerAlaLeuAspPheIleLysLeuAspPheTy 617
|||||
2077 CATTTGCCCAATTCTGCACACTGCTGACTTCATTAATTCGACATTGTA 2126
|||||
617 rGlyGlyAlaMetAlaSerProGluLysAlaIleGluAspThrGlyGlyI 634
|||||
2127 TGGGGAGCTATGGCTTCATGGGAAAGGCTGCAGAGACACAGGTGGA 2176
|||||
634 lHisMetGlnGluAlaProGluThrTyrIleProSerArgAlaValSer 650
|||||
2177 TCCACATGGAAGAGGCCCCCAAAACCTACATTCACAGCAGGCTGTATCT 2226
|||||
651 LeuPhePheAsnThrLysGlnGluPheArgThrLeuGluValThrLeuAr 667
|||||
2227 TTGTCTTCAACTGGAAGCAGGAATTCAGACCTGTGAGGTCTCACCTCC 2276
|||||
667 gaSPheSerLysLeuAsnLysGlnAspIleThrTyrLeuGlyLysIleP 684
|||||
2277 GCATTTACGAACTTGATATAGCAAGATATACATATCTGGGAAATAT 2326
|||||
684 heSerSerAlaThrSerLeuArgLeuGlnIleLysArgCysAlaGlyVal 700
|||||
2327 TCAGCTCTGCCACAGCCTCAGGCTGCAATTAAGATGTGCTGGGTG 2376
|||||
701 AlAGlySerLeuSerLeuValLeuSerThrCysLysAsnIleTyrSerLe 717
|||||
2377 GCTGGAAAGCCTCAGTTGGTCTCAGACCTGTAAAGCACTTTATTTCTCT 2426
|||||
717 uMetValGluAlaSerProLeuThrIleGluAspGluArgHisIleThrS 734
|||||
2427 CATGTGGAAAGCCAGTCCCTCAACATAGAAAGTGAAGGACACATACAT 2476
|||||
734 eValThrAsnLeuLysThrLeuSerIleHisAspLeuGlnAsnGlnArg 750
|||||
2477 CTGTAAACAAACCTGAACCTTGAGTATTCATGACCTACAGATCAACGG 2526
|||||
751 LeuProGlyGlyLeuThrAspSerLeuGlyAsnLeuLysAsnLeuThrLy 767
|||||
2527 CTGCGGGGTGCTGACTGACACTGGGTGAACCTTAAGAAACCTTCAAA 2576
|||||
767 sleuIleMetAspAsnIleLysMetAsnGluGluAspAlaIleLysLeu 784
|||||
2577 GCTCATATATGATTAACATTAAGATGAATGAAGAAAGTACTATAAACTGA 2626
|||||
784 lAlaGlnGlyLeuLysAsnLeuLysLysMetCysLeuPheHisLeuThrHis 800
|||||
2627 CTGAAGGCCCTGAACAAACCTGAAGAGATGTGTTATTTCAATTGACCCAC 2676
|||||
801 LeuSerAspIleGlyGlyMetAspTyrIleValAllySerLeuSerSe 817
|||||
2677 TTGTCTGACATTTGGAAGGGAATGATTAATAGTAAAGTCTCTGTCAAG 2726
|||||
817 rGluProCysAspLeuGluGluIleGlnLeuValIserCysCysLeuSerA 834
|||||
2727 TGAACCTGTGACCTTGAAAGAAATTCATTAATAGTCTGCTGCTGTGTG 2776
|||||
834 lAsnAlaValLysIleLeuAlaGlnAsnLeuHisAsnLeuValLysLeu 850
|||||

```

2777 CAATATCAGGAAATATCCTAGCTCAGAAATTCACATTTGGTCAAACTG 2826

851 Set11leuAspIeuSerGluAsnTYrLeuGluLysAspGlyAsnGluAl 867

2827 AGCATTCCTGATTATTCAGAAATTCCTGGAAAAAGATGGAATGAGC 2876

867 albeuHISglLeuIleAspArgMetAsnValLeuGluGlnLeuThrAla 884

2877 TCTTCATTGAACCTGATCGACAGGATGAACCTGCTAGAACACCTACCGCAC 2926

884 euMetLeuProTrrGlyCysAspValGlnGlySerLeuSerSerLeu 900

2927 TGTATGCTGCCCTGGGGCTGGAGCTGCAAGGACGCTGAGCACCTGTTG 2976

901 LysHISLeuGluGlnValProGlnLeuValLysLeuGlyLeuLysAsnTr 917

2977 AAACATTTGGAGAGGGTCCACCAACTGCTCAAGCTTGCGTTGAAAAACTG 3026

917 ParGleuThrAspThrGluIleArgIleLeuGlyAlaPhePheGlyLysA 934

3027 GACACCTCACGATACAGAGATTGAAATTTAGTGCATTTTGGAGAGAGA 3076

934 snProLeuLysAsnPhenGlnGlnLeuAsnLeuAlaGlyAsnArgValSer 950

3077 ACCCTCTGAAAAACTTCACAGCAGCTGAAATTTGGCGGGAATTCGTGGAGC 3126

951 SerAspGlyTrrLeuAlaPheMetGlyValPheGluAsnLeuLysGlnLe 967

3127 AGGATGGATGGCTTCCTCATGGGTGTATTGAGAAATCTTTAAAGCAATT 3176

967 uValPhePheAspPheSerThrLysGluPheLeuProAspProAlaLeu 984

3177 AGGTTTTTTGACTTTAGTACTTAAGAAATTTCTACCTGATCCAGCATTAG 3226

984 alArgLysLeuSerGlnValLeuSerLysLeuThrPheLeuGlnGluAla 1000

3227 TCAGAAACTTACCCAAAGTGTATCCAAAGTTACTTTTCGCAAGAAAGCT 3276

1001 ArgLeuValGlyTrrGlnPheAspAspAspAspLeuSerValIleHrgI 1017

3277 AGCGTTGTTGGGTGGCAATTTGATGATGATGATCAGTGTATTTCACAG 3326

1017 yAlaPheLysLeuValThrAla 1024

3327 TGCCTTTAAACTAGTAACCTGCT 3348

seq_name: gb_hcg.AC010968

seq_documentation_block:

LOCUS AC010968 160583 bp DNA HTG 18-AUG-2000

DEFINITION Homo sapiens chromosome 2 clone RP11-9302, WORKING DRAFT SEQUENCE, 11 unordered pieces.

ACCESSION AC010968

VERSION AC010968.5 GI:9845170

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 160583)

AUTHORS Waterston,R.H.

TITLE The sequence of Homo sapiens clone unpublished

REFERENCE 2 (bases 1 to 160583)

AUTHORS Waterston,R.H.

JOURNAL Direct Submission

Submitted (28-SEP-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis MO 63108, USA

COMMENT On Aug 18, 2000 this sequence version replaced gi:8439959.

```

Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
-----
Project Information
-----
Center project name: H_NH0093002
Summary Statistics
-----
Sequencing vector: M13; 59k
Chemistry: Dye-primer ET; 47% of reads
Chemistry: Dye-terminator Big Dye; 53% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: bases at least Q40
Consensus quality: bases at least Q30
Consensus quality: bases at least Q20
Insert size: 147000; agarose-fp
Insert size: 159583; sum-of-contigs
Quality coverage: 6.64 in Q20 bases; agarose-fp
Quality coverage: 6.38 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1      1080: contig of 1080 bp in length
*
1081      1180: gap of unknown length
*
1181      2476: contig of 1296 bp in length
*
2477      2576: gap of unknown length
*
2577      5035: contig of 2459 bp in length
*
5036      5135: gap of unknown length
*
5136      9606: contig of 4471 bp in length
*
9607      17078: gap of unknown length
*
17079      17178: contig of 7372 bp in length
*
17179      27158: gap of unknown length
*
27159      27258: contig of 9980 bp in length
*
27259      45137: gap of unknown length
*
45138      45237: gap of 17879 bp in length
*
45237      65523: contig of 20285 bp in length
*
65523      65623: gap of unknown length
*
65623      91498: contig of 25876 bp in length
*
91499      91598: gap of unknown length
*
91599      116835: contig of 25237 bp in length
*
116836      116935: gap of unknown length
*
116936      160583: contig of 43648 bp in length.
*
Location/Qualifiers
1.160583
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/clone="RP11-9302"
1.1080
/misc_feature      /note="assembly_name:Contig9"
1181.2476
/misc_feature      /note="assembly_name:Contig12"
2577.5035
/misc_feature      /note="assembly_name:Contig13"
5136.9606
/misc_feature      /note="assembly_name:Contig14"
9707.17078
/misc_feature      /note="assembly_name:Contig15"
17179.27158
/misc_feature      /note="assembly_name:Contig16"
27259.45137
/misc_feature      /note="assembly_name:Contig17"
45238.65522
/misc_feature      /note="assembly_name:Contig18"
vector_side:right"
65623.91498
/misc_feature      /note="assembly_name:Contig19"

```

misc_feature 91599..116835
/note="assembly_name:Config20
clone_end:sp6
vector_side:right"
misc_feature 116936..160583
/note="assembly_name:Config21"
BASE COUNT 44349 a 33843 c 33703 g 47641 t 1047 others
ORIGIN

alignment_scores:
Quality: 3733.00 Length: 1067
Ratio: 4.854 Gaps: 4
Percent Similarity: 72.071 Percent Identity: 70.947

alignment_block:

S-09-697-089-2 x AC010968 ..

align seg 1/1 to: AC010968 from: 1 to: 160583

```
1 MetAspHellelysaspasnserrargalaileuileglnarymetgyme 17
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
48109 GTCGAATTTCTAAAGACAAATACCCGACCTTATTCAAAGAAATGGGAAT 48158
17 tthrvalilleysglnlethrapsaspleuphevaltrpasnvalleua 34
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
48159 GACTGTTATAAGCAAAATTCACAGATGCTATTGTATGGAATGTTCTGA 48208
34 snarpglugluvalasnillelecyssglsulysvalglulinaspala 50
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
48209 ATGCGCAAGAGTAACATCATTTGCTGCGAGAGGTGACAGCAAGATGCT 48258
51 Alarargglyllellehsmelleleuylslysglyserglusercysas 67
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
48259 GCTAGAGGATCATTCACATGATTTTGAAAAGGTTTCAGAGTCCGTATA 48308
67 nleupheleuylserserleuylsglutrpsantryproleupheglnaspl 84
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
48309 CCCTTTCTTAAATCCCTTAAGAGAGTGAACATACTCTATTTTCAGAGACT 48358
84 euasnnglyln..... 87
  ::::::::::::::
48359 TGAATGACACAAGTAAGTATCAGTAGTTTGCATCTGTTCTGCAAGTA 48408
87 ..... 87
  ::::::::::::::
48409 GAATAGACCAACAACAAGATCATTTCCCATGGCAGAGCCTCTGCTTT 48458
87 ..... 87
  ::::::::::::::
48459 TCCTTCTTCATATACACTGCGCAGTGGCTTTGATACCTGCTAAGAAA 48508
87 ..... 87
  ::::::::::::::
48509 AGCTAGATGATTTTAGAAAACTCATAGTATCATTAATGATGATGAAG 48558
87 ..... 87
  ::::::::::::::
48559 CCAAAACAATAACCCCAAGGCCAGGAATAAATGATGAGTATTGTCTCA 48608
87 ..... 87
  ::::::::::::::
48609 CTGCAAGATAAAACTGAGATAAAGTCACTAGTTAAGTTCACAGTCCAG 48658
87 ..... 87
  ::::::::::::::
48659 CAGTGTCTTCAGCAATCCCTGGTGGATGAGATCTACTATGAGAGAGA 48708
87 ..... 87
  ::::::::::::::
48709 GGGCCAGCTACTACTACTCATGTCACAGGCTCTGCTGTAAGACAGCC 48758
87 ..... 87
```

```
48759 TGACTGACAGAACCATGATTCATGTTTCATAAATATTGTTTTATT 48808
87 ..... 87
  ::::::::::::::
48809 TTTTAAATAGTAATAATAGGTCATGACAGATTTTACTCAGAGTGAAGGCC 48858
87 ..... 87
  ::::::::::::::
48859 TACTATGCTGACAATTATCCCTAACCATCTTCTCTGCCAAATCTTTAC 48908
87 ..... 87
  ::::::::::::::
48909 TTTCTTAACCAAAAAACAGAGTCCTCCCTTTTCCAAACGACGTGTATAA 48958
87 ..... 87
  ::::::::::::::
48959 GATTGGTGTGGAAAGAGAGAAACAGAAACAAAAAGGGAGTGGGGTA 49008
87 ..... 87
  ::::::::::::::
49009 AAGGATGCTGGGAAAGGGGAGCAGACAGAACCTTTGTGCCAGCTCAG 49058
87 ..... 87
  ::::::::::::::
49059 AACTGGGAAGACATTTACATTCACCAATGGCAGATTTCAGGCTCAGCC 49108
87 ..... 87
  ::::::::::::::
49109 CCCCCTTCATCTCTGAACTGAGTGTGAGAGACACAGAAATGGGT 49158
88 ..... 87
  ::::::::::::::
49159 CTTCCTTAACATCACCCTTTCTTAATAGTCTTTTTCACACACATCAG 49208
95 luelyaspleuasaspaleualaglnaspleuylsaspheutyrhsthr 111
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
49209 AAGGAGCTTGAGAGATTTGGCTCAGAGATTTAAAGACTGTGACATPAC 49258
112 ProSerPheleuasnbethyrProleuglygluaspllealeph 128
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
49259 CCATCTTTCTGAACCTTATCCCTTGTGGAAGATATTGACATTTATTTT 49308
128 easnleuylserserthpethrgluprovalleutprarglysasglnh 145
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
49309 TAACTTGAAAGACCTTCACAGAACCTGCTGTGAGAGAAAGGACACAC 49358
145 lshshisargvalgluglnleutthreuasnglyleuileuglnalaleu 161
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
49359 ACCATCACCGCGTGAGACAGCTGACCTGAATGGCCTCGCAGGCTCTT 49408
162 GluSerProCyslleilegluglyluserglylysglylerserthle 178
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
49409 CAGAGCCCTCGCATCATTTGAAGGGGAATCGGCAAGGCAAGTCCACTCT 49458
178 uleuglnargllealemetleutrrpglyserglylyslyslealeut 195
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
49459 GCTGCAGCCGATTCATGCTGCGGCTCCGGAAGGTGCAAGGCTCTGA 49508
195 hrlyspheuylsphaValphepheleuargleuserargalaglinglely 211
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
49509 CCAAGTTCAAAATTCGCTCTTCTTCCTCGCTCAGAGGGGCCAGGGTGGA 49558
212 leuphegluThrleucysaspglnleuAsplleprogllythrlleat 228
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
49559 CTTTGGAAACCTCTGTGATCAATCTCGGATATACCTGGCACAATCAG 49608
228 glysglnThrPheMetAlaMetleuylsleuarglnargvalleup 245
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
49609 GAAGCAGACATTCATGCGCTGCTGAGAGCTGCGGACGAGGGTCTTT 49658
245 heleuleuaspglytyrasngluPheylsProglinsncysprogluile 261
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
```

49659 TCCTCTTGATGGCTACAAATGATTCAGACCCAGAACCTGCCAGAAATTC 49708
 262 GluAlaLeuLeuLeuGluAsnHisArgPheLysAsnMetValIleValIle 278
 49709 GAAGCCCTGATTAAGAAACACCGCTTCAGAACATGTGTATGTCTGC 49758
 278 rThrThrThrGlyCysLeuArgHisIleArgGlnPheGlyAlaLeuThrA 295
 49759 CACTACCACTGAGTGCCTGAGGCACATACGGCAGTTGGTCCCTGACTG 49808
 295 IagIValGlyAspMetThrGluAspSerIagIAlaLeuLeuLeuArgGlu 311
 49809 CTGAGGTGGGGATATGACAGAAAGAGCGCCAGGCTCTCATTCGGAGA 49858
 312 ValLeuIleuGlyGluLeuAlaGluGlyLeuLeuLeuGlnIleGlnLys 328
 49859 GTGCTGATCAAGAGCTGCTGAGAGCTTGTGCTCCAAATTCAGAAATC 49908
 328 rArgCysLeuArgAsnLeuMetLysThrProLeuPheValValIleThrC 345
 49909 CAGGTGCTTGAGAAATCTCATAGAACCCCTCTTGTGGTCATCAGACT 49958
 345 ySAIleIleGlnMetGlyGluSerGluPheHisSerHisThrGlnThr 361
 49959 GTGCAATCCAGATGGGTGAAGTAGATTCACCTCACACACAAACACAG 50008
 362 LeuPheHisThrPheThrAspLeuLeuIleGlnLysAsnLysHisLysH 378
 50009 CTGTTCCATACCTCTATGATGTGTGATACGAAACAAACACACAAACA 50058
 378 sLysGlyValAlaAlaSerAspPheIleArgSerLeuAspHisCysGlyA 395
 50059 TAAAGTGTGGCTGCAGATGATTCATTCGAGCCCTGAGCCACTGTGAG 50108
 395 sPLeuAlaLeuGluGlyValIlePheSerHisLysPheAspPheGluLeuG 411
 50109 ACCTAGCTCTGAGAGGTGTGTCTCCACAAAGTTGATTCGACATGGAG 50158
 412 AspValSerSerValAsnGluAspValLeuLeuThrThrGlyLeuLeu 428
 50159 GATGTCTCCAGCGTGAATGAGATGCTGCTGACAACTGGGCTCCTCG 50208
 428 sLysThrThrAlaGlnArgPheLysProLysThrLysPhePheHisLys 445
 50209 TAAATATACAGCTCAAGGTTCAAGCCAAAGTAAATTCCTTTCACAGT 50258
 445 ePheGlnGluThrThrAlaGlyArgArgLeuSerSerLeuLeuThrSer 461
 50259 CATTCAGAGAGACACAGCAGACGAACTCAGCAGATTATTGACGCTCT 50308
 462 HisGluProGluGluValThrLysGlyAsnGlyThrLeuGlnLysMet 478
 50309 CATGAGCCAGAGAGGTGACCAAGGGGAATGTTACTTCAGAGAAATGCT 50358
 478 IserLleSerAspIleThrSerThrThrLysSerSerLeuLeuArgThr 495
 50359 TTCATTTTCGGCATTACATCCATTATGACGCTGCTCCGCTACACT 50408
 495 ySGIleSerSerValGluAlaThrArgAlaValMetLysHisLeuAla 511
 50409 GTGGGTCTATCTGTGAGACGACAGGCTGTTATATAGACACTCCAGCA 50458
 512 ValThrGlnHisGlyCysLeuLeuGlyLeuSerIleAlaLysArgPro 528
 50459 GTGTATCAACAGCGCTGCTCTGAGACTTTCATCGCCCAAGAGCCCT 50508
 528 uTPArgGlnGluSerLeuGlnSerValLysAsnThrThrGluGlnLui 545
 50509 CTGGAGACAGGAATCTTGCAAAAGTGAACCAACCACTGAGCAAGAA 50558
 545 IeLeuLeuAlaIleAsnIleAsnSerPheValGluCysGlyLleHisLeu 561
 50559 TTCTTAAAGCCCTAATCATCTCTTTGTAGAGTGTGGCATTCATTTA 50608

562 TyrGlnGluSerThrSerLysSerAlaLeuSerGlnGluPheGluAlaP 578
 50609 TATCAAGAGAGATCATCAANTCAGCCCTGAGCCAGAAATTTGAAGCTTT 50658
 578 ePheGlnGlyLysSerLeuThrIleAsnSerGlyAsnIleProAspThr 595
 50659 CTTTCACAGTAAACCTTATATATCATCACTCAGGAGAACATCCCGCTACT 50708
 595 euPheAspPhePheGlnHisLeuProAsnGlyAlaSerAlaLeuAspPhe 611
 50709 TATTTGACTTCTTTACACATTTGCCAATGTGCAAGTGTCTGAGCTTC 50758
 612 IleLysLeuAspPheThrGlyGlyAlaMetAlaSerTrpGluLysAla 628
 50759 ATTAACTGAGACTTTTATGGGGAGCTATGGCTTCATGGAAAGAGCTGC 50808
 628 aGluAspThrGlyGlyIleHisMetGluGluAlaProGluThrThrIleP 645
 50809 AGAAGACACAGGTGGAAATCCACATGGAAGAGCCCAAGAACTACATTC 50858
 645 rSerArgAlaValSerLeuPheAsnTrpLysGlnGluPheArgThr 661
 50859 CCAGCAGGCGCTGATCTTTGTTCTTCACTGAGAACAGGAATTCAGACT 50908
 662 LeuGluValThrLeuArgAspPheSerLysLeuAsnLysGlnAspIle 678
 50909 CTGAGGTGTACACTCCGGGATTTACAGAAATGATTAAGCAAGATACAG 50958
 678 rTyrLeuGlnGlyLysIlePheSerSerAlaThrSerLeuArgLeuGlnIle 695
 50959 ATATCTGGGGAATATTCAGCTCTGCCAAGCCCTCAGGCTGCAGAAATA 51008
 695 ySArGlyAlaGlyValAlaGlySerLeuSerLeuValLeuSerThrCys 711
 51009 AGAGTGTGCTGCTGTGCTGAGACCTCAGTTGCTGCTCAGACCTGCT 51058
 712 LysAsnIleThrSerLeuMetValGluAlaSerProLeuThrIleGluAs 728
 51059 AAGACATTTATTTCTCTCATGTGGAGACCAAGTCCCTCACATGAGAGA 51108
 728 pGluArgHisIleThrSerValThrAsnLeuLysThrLeuSerIleHisA 745
 51109 TGAGAGGCACATCATCTGTAAACAACTGAAACCTTGATATTCATG 51158
 745 sPLeuGlnAsnGlnArgLeuProGlyGlyLeuThrAspSerLeuGlyAsn 761
 51159 ACCTACAGATCAACGCTGCCGGGT..... 51184
 762 LeuLysAsnLeuThrLysLeuIleMetAspAsnIleLysMetAsn.... 776
 51185 ATTGTAATATACAGAGTGTGCTGCTTGTTCACCTTAAAAAAAATAACAG 51234
 777GluGluAspAlaIleLysLeuAlaGluGly 786
 51235 TGTATAAATTTGGAAAGAGAGAGAGACTTATTTCTTCTTAAAGGCT 51283

seq_name: gb_CNS01DS3

seq_documentation_block:

LOCUS CNS01DS3 138909 bp DNA PRI 02-MAR-2000

DEFINITION BAC sequence from the SP4 candidate region at 2p21-2p22 BAC 164M19

complete sequence.

ACCESSION AL121653

VERSION AL121653.2 GI:7159616

KEYWORDS HTG; HTGS_DRAFT; SP4 genomic DNA interval.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 138909)

Hazan, J., Fonknechten, N., Mave, D., Paternotte, C., Samson, D.,

Artiguenave, F., Davoine, C.S., Cruaud, C., Durf, A., Wincker, P.,
 Brotier, P., Catolico, L., Barbe, Y., Burzender, J.M.,
 Prud'Homme, J.F., Brice, A., Fontaine, B., Hellig, R. and
 Weissenbach, J.

TITLE Spastin, a novel AAA protein, is altered in the most frequent form
 of autosomal dominant spastic paraplegia
 Nat. Genet. (1999) In press
 2 (bases 1 to 138909)

JOURNAL Direct Submission
 REFERENCE Submitted (29-FEB-2000) to the EMBL/GenBank/DBJ databases
 COMMENT On Mar 6, 2000 this sequence version replaced g1:6002386.
 FEATURES Location/Qualifiers

Source

1. 138909
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /clone_11b="C17B_978_SKB"
 /clone="164M19"
 /clone="164M19"

BASE COUNT 39243 a 28424 c 29121 g 42121 t
 ORIGIN

alignment_scores:

Quality: 3733.00 Length: 1067
 Ratio: 4.854 Gaps: 4
 Percent Similarity: 72.071 Percent Identity: 70.947

alignment_block:

us-09-697-089-2 x CNS01DS3/rev ..

Align seg 1/1 to reverse of: CNS01DS3 from: 1 to: 138909

```

1 MetAspHeIleIySaSpAsnSerArAlaLeuIleGlnArgMetGlyme 17
  ::::::::::::::::::::::::::::::::::::::::::::::::::::
76388 GTGAAATTCTAAAGACAAATAGCCGACCCCTATTCAAAAGATGGGAAT 76339
17 tThValIleLysGlnIleThraSpAspLeuPheValTrpAsnValLeuA 34
  ::::::::::::::::::::::::::::::::::::::::::::::::::::
76338 GACTGTTATTAAGCAAAATACAGATGACCTATTGTTGATGGAATGTTCTGA 76289
34 snArgGluGluValAsnIleIleCysCysGluLysValGluGlnAspAla 50
  ::::::::::::::::::::::::::::::::::::::::::::::::::::
76288 ATGCGAAGAAGTAAACATCATTTGCTGCGAAGAAGTGAAGCATGCT 76239
51 AlaArgGlyIleIleHisMetIleLeuLysLysGlySerLysCysAs 67
  ::::::::::::::::::::::::::::::::::::::::::::::::::::
76238 GCTAGAGGATCATTCACATGATTTTGAAAAAGGTTCAAGATCCTGTAA 76189
67 nLeuPheLeuLysSerLeuLysGluTrpAsnTyrrProLeuPheGlnAspL 84
  ::::::::::::::::::::::::::::::::::::::::::::::::::::
76188 CCTCTTCTTAATCCCTTAAGAGTGAACATCTCTTAATTCAGAGACT 76139
84 euAsnGlyLysIn..... 87
  ::::::::::::::
76138 TGAATGACAAAGTAAGTATCAGTTAGTTTGCATCTGTTTCGCAAGTA 76089
87 ..... 87
76088 GAATAGACCAACAAGATCATCTTCCCATGCGAGAGCCCTGCTTT 76039
87 ..... 87
76038 TCCTTCTTCATATCACCTGCGACGTGGCCTTGATACCTGCTAAGGAA 75989
87 ..... 87
75988 AGCTTAGATGATTTTAGAAAAACTATAGTATCAATTGATGCTGAAG 75939
87 ..... 87
75938 CCAAAACAATAACCCCAAGCCAGGAATAATGATTGACTAATTGCTCA 75889

```

```

87 ..... 87
75888 CTGCAAGATAAAACTGAGATAAGTCACTAGTTAAGTTCACATGCGAC 75839
87 ..... 87
75838 CAGTGGTTCTCAGCATTCCTGTTGGATGAGATCTACATATGGGAGAGA 75789
87 ..... 87
75788 GGCCACAGCTACTACTACTCATGTCACAGCCCTCTGTAAAGACAGACC 75739
87 ..... 87
75738 TGACTGACACAGAACCATGATTCTATGTTTCATTAATAATATTGTTTAT 75689
87 ..... 87
75688 TTTTAAATAGTAATAAGTCAATGACAGATTTTACTCAGAGTGAAGGCC 75639
87 ..... 87
75638 TACTATGATGACAATTATCCCTAACCATCTTCTGCGCAATTCCTTAC 75589
87 ..... 87
75588 TTTCCTTAACCCAAAACAGCTCTCCCTTTTCCACACAGCTGTATTA 75539
87 ..... 87
75538 GATTGGTTGTGAAGAGAGAAACAGAAAGCAAAAGGGAGTGGGGTA 75489
87 ..... 87
75488 AAGATCTGGGGAAAGGGGAGCAGACAGAACCCCTTGTGCCAGCTCAG 75439
87 ..... 87
75438 AACTGGAAAGACATTTACTTACCAATGCGAGATTTCAGCCTCACCC 75389
87 ..... 87
75388 CCCCCTCATCTCTGAACTAGTTTTGAGAGACACAGAAATGGGT 75339
87 ..... 87
75338 CTTCCTTAACATCACCCTCTCTCTAATAGGTCTTTTCATCAGACATCAG 75289
95 LuGlyAspLeuAspAspLeuAlaGlnAspLeuLysAspLeuTyrrHisThr 111
  ::::::::::::::::::::::::::::::::::::::::::::::::::::
75288 AAGGAGACTTGGACGATTTGGCTCAGAGATTTAAAGACTTGTACCATACC 75239
112 ProSerPheLeuAsnPheTyrrProLeuGlyGluAspIleAspIleIlePh 128
  ::::::::::::::::::::::::::::::::::::::::::::::::::::
75238 CCAATCTTTCTGAACCTTTATCCCTTGGGAGAGATATTAGCATTTATTT 75189
128 euAsnLeuLysSerThrPheThrGluProValLeuTrpArgLysAspGlnH 145
  ::::::::::::::::::::::::::::::::::::::::::::::::::::
75188 TAACTGAAAAGCACCTTCACAGAACCTGCTGCTGAGAGAAAGACCAAC 75139
145 IShHisArgValGluGlnLeuThrLeuAsnGlyLeuGlnAlaLeu 161
  ::::::::::::::::::::::::::::::::::::::::::::::::::::
75138 ACCATCACCGCGTGGAGCAGCTGACCTGATGCGCTCTGCGAGGCTCTT 75089
162 GlnSerProCysIleIleGluGlyLysSerGlyLysGlyLysSerThrLe 178
  ::::::::::::::::::::::::::::::::::::::::::::::::::::
75088 CAGAGCCCTCGATCATTTGAGGGGAATCGGCAAGGCAAGTCCACTCT 75039
178 uLeuGlnArgIleAlaMetLeuTrpLysSerGlyLysCysLysAlaLeu 195
  ::::::::::::::::::::::::::::::::::::::::::::::::::::
75038 GCTGACAGCAATTCATGCTCTGGGGCTCCGGAAGAGTGAAGGCTCTGA 74989
195 hTrpSerPheValPhePheLeuArgLeuSerArgAlaGlnGly 211

```

```

|||||
74988 CCAAGTCAAAATGGTCTTCTTCCTCCCTCAGAGGAGGCGGAGTGA 74939
212 LeuphegluThrLeucysAspGlnLeuLeuAspIleProGlyThrIleArg 228
|||||
74938 CTTTGTGAACCCCTGTGTGATCAACTCCTGGATATACCTGGACAAATCAG 74889
228 gLysGlnThrPheMetIleMetLeuLeuLysLeuArgGlnArgValLeuP 245
74888 GAAGGACGACATTCATGCGCATGCTCTCAAGCTGCGCAGAGAGGTTCTTT 74839
245 helLeuLeuAspGlyTyrAsnGlnPheLysProGlnAsnGlyProGluLe 261
|||||
74838 TCCTCTGTGATGGCTACAAATGATTCAGCCCAAGCTGCCCCAGAAATTC 74789
262 GluAlaLeuIleLysGlnAsnHisArgPheLysAsnMetValIleValThr 278
74788 GAAGCCCTGATTAAGAAACCACCGCTTCAAGAACATGCTCATGCTCAG 74739
278 rThrThrThrGlyCysLeuArgHisIleArgGlnPheGlyAlaLeuThrA 295
74738 CACTACCACTAGTGTCTGAGGCACATACGGCAGTTGGTGGCTGACTG 74689
295 IagIuValGlyAspMetThrGluAspSerAlaGlnAlaLeuIleArgGlu 311
74688 CTGAGGTGGGGATATGACAGAGACAGCCAGGCTCTCATCCGAGAA 74639
312 ValLeuIleLysGlnLeuAlaGlnGlyLeuLeuGlnIleGlnLysSe 328
74638 GTGCTGATCAAGAGACTGCTGAAAGCTTGTGCTCCAATTCAGAAATC 74589
328 rArgCysLeuArgAsnLeuMetLysThrProLeuPheValValIleThrC 345
74588 CAGGTGCTTGAGGAATCATGAAAGACCCCTCTTGTGGTCAATCACTT 74539
345 ySAIleGlnMetGlyLysSerGlnPheHisSerHisThrGlnThrThr 361
74538 GTGCATTCAGATGGGTGTAAGTGAATTCACCTCAGACACAAACAACG 74489
362 LeuPheHisThrPheTyrAspLeuLeuIleGlnLysAsnLysHisLysH 378
74488 CTGTTCATACCTCTTATGATCTGTGATACGAAACAAACAAACA 74439
378 sLysGlyValAlaIleAspPheIleArgSerLeuAspHisCysGlyA 395
74438 TAAAGCTGTGCTGCACGTGATTCATTCGGCCTGAGCCACTGTGAG 74389
395 sPLeuAlaLeuGlnGlyValPheSerHisLysPheAspPheGlnLeuGln 411
74388 ACCTAGCTCTGGAGGGTGTGTTCTCCCAAGTTTGATTTGCAACTGAG 74339
412 AspValSerSerValAsnGlnAspValLeuLeuThrThrGlyLeuLeuCy 428
74338 GATGTGTCACAGCGTAATGAGGATGCTGCTGCAACTGGGCTCTCTG 74289
428 sLysTyrThrAlaGlnArgPheLysProLysTyrLysPhePheHisLys 445
74288 TAAATATACACTCAAGGTTCAAGCCAAAGTATTAATCTTTCACAAGT 74239
445 erPheGlnGluTyrThrAlaGlyArgArgLysSerLeuLeuThrSer 461
74238 CATTCACAGAGTACACAGCAGAGACAGACTCAGAGTTTATGACGCT 74189
462 HisGluProGlnGluValThrLysGlnLysAsnGlyTyrLeuGlnLysMe 478
74188 CATGAGCCAGAGAGGTGACCAAGGGGAATGTTACTTGCAAGAAATGCT 74139
478 IserIleSerAspIleThrSerThrTyrSerSerLeuLeuArgTyrThrC 495
74138 TTCCATTTCCGACATTACATCCACTTATAGACGCTGCTCGGTACACCT 74089
495 ySGlySerSerValGluAlaThrArgAlaValMetLysHisLeuAla 511
|||||
74088 GTGGGTCAATCTGTGGAGCCACCAAGGCTGTATTGAAAGCACTGCGACA 74039
512 ValTyrGlnHisGlyCysLeuLeuGlyLeuSerIleAlaLysArgProLe 528
74038 GTGTATCAACAGCGCTGCTTCTGCGACTTTCATCGCCAGAGAGGCTCT 73989
528 uTPArgGlnGlnSerLeuGlnSerValLysAsnThrThrGlnGlnGlu 545
73988 CTGGAGACAGGAATCTTGGCAAAAGTGAATAAACACCACTGAGCAAGAA 73939
545 IeLeuLysAlaIleAsnLysSerPheValGluCysGlyIleHisLeu 561
73938 TTCTGAAGCCATTAACATCAATTCCTTTGTAGAGTGTGGCATCATTTA 73889
562 TyrGlnGlnSerThrSerLysSerAlaLeuSerGlnGlnPheGluAlaPh 578
73888 TATCAAGAGAGTACATCCAAATCAAGCCCTGAGCCCAAGAAATTTGAAGCTT 73839
578 erPheGlnLysSerLeuTyrIleAsnSerGlyAsnIleProAspTyrL 595
73838 CTTTCAAGGTAAACCTATATATCACTCAGGGAACATCCCGATTA 73789
595 euPheAspPhePheGlnHisLeuProAsnGlyAlaSerAlaLeuAspPhe 611
73788 TATTTGACTTCTTGTGAACATTTGCCCAATTTGTGCAAGTGTCCCTGACTTC 73739
612 IleLysLeuAspPheTyrGlyValAlaMetAlaSerTPGlyLysAlaAl 628
73738 ATTAACTGGAATTTTATGGGGAGCTATGGCTTCATGTGAAAGGCTGC 73689
628 gLysAspThrGlyLysIleHisMetGluGluAlaProGluThrTyrIleP 645
73688 AGAAGACACAGGTGGAATCCACATGGAAGAGGCCCAAGAACCTACATTC 73639
645 roSerArgAlaValSerLeuPhePheAsnTrpLysGlnGluPheArgThr 661
73638 CCAGCAGGCTGTATCTTTGTTCTTCACTGGAAGCAGGAATTCAGGACT 73589
662 LeuGluValThrLeuArgAspPheSerLysLeuAsnLysGlnAspIleTh 678
73588 CTGAGGTGCACACCTCGGGATTTGAGCAAGTTGAATAGCAAGATATCAG 73539
678 rTyrLeuGlnLysIlePheSerSerAlaThrSerLeuArgLeuGlnIleL 695
73538 ATATCTGGCGAAATATTCACCTCTGCCCAAGCCTCAGCCTGCAAAATTA 73489
695 ySArGlyAlaGlyValAlaGlySerLeuSerLeuValLeuSerThrCys 711
73488 AGAGATGTGCTGTGTGCTGGAAGCTTCAGTTTGTGCTCAGCACTGT 73439
712 LysAsnIleTyrSerLeuMetValGluAlaSerProLeuThrIleGluAs 728
73438 AAGAACAATTTATTTCTCATGTGTGAAGCCAGTCCCTCACCATAGAAAGA 73389
728 pGluArgHisIleThrSerValThrAsnLeuLysThrLeuSerIleHisA 745
73388 TGAGAGGCACATCATCTGTACAAACCTGAAACCTTGAGTATTCATG 73339
745 sPLeuGlnAsnGlnArgLeuProGlyGlyLeuThrAspSerLeuGlyAsn 761
73338 ACCTACAGAAATCAAGGCTGCCGGGT..... 73313
762 LeuLysAsnLeuThrLysLeuIleMetAspAsnIleLysMetAsn.... 776
73312 APTGAAATATACAGCTTGTGCTTGTTCCTTAACCTTAAAAAATAATACAG 73263
777 .....GluGluAspAlaIleLysLeuAlaGluGly 786
73262 TGTATTAATTTGAAAAAGAGAGAGAGACTTATTTCTCTTAAAGGT 73214
seq_name: gb_pr.IR2005417
seq_documentation_block:

```

LOCUS IR2005417 1355 bp mRNA PRI 16-JUL-2000
DEFINITION Homo sapiens mRNA full length insert cDNA clone EUROIMAGE 2005417.
ACCESSION AL389934
VERSION AL389934.1 GI:9367839
KEYWORDS FLI_CDNA.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 1355)
AUTHORS Auffray,C., Ansoorge,W., Ballabio,A., Estivill,X., Gibson,K.,
Lehach,H., Pouslka,A. and Lundberg,J.
The European IMAGE consortium for integrated Molecular analysis of
human gene transcripts
Unpublished
2 (bases 1 to 1355)
JOURNAL Pluvinet,R., Estivill,X., Escarceller,M. and Sumoy,L.
REFERENCE Direct Submission
AUTHORS Submitted (15-JUL-2000) Dept. Genetica Molecular, Institut de
JOURNAL Recerca Oncologica (IRO), Hospital Duran i Reynals, Av. Gran Via
s/n Km 2,7 L'Hospitalet de Llobregat, 08907 Barcelona, Catalunya,
SPAIN. Tel: ++34-93-260-7775 Fax: ++34-93-260-7776 WWW site:
http://www.iro.es e-mail enquiries: lsumoy@iro.es
EURO-IMAGE Consortium Contact: Auffray C
CNRS UPR 420 - Genetique Molculaire et Biologie du Developpement
IFR 1221 - Rue Guy Moquet 19, Batiment G - Bp 8
94801 Villejuif Cedex, FRANCE
Tel: ++33-1-49 58 34 98
Fax: ++33-1-49 58 35 09
e-mail: auffray@infobiogen.fr
This clone is available royalty-free through IMAGE Consortium
Distributors.
IMPORTANT: This sequence represents the full insert of this IMAGE
cDNA clone. No attempt has been made to verify whether this
corresponds to the full-length of the original mRNA from which it
was derived.

FEATURES
source Location/Qualifiers
1..1355
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/clone="IMAGE cDNA clone 2005417"
/clone_1ib="NCI_CGAP_Paul"
/map="2p21-p22"
1..1321
/codon_start=2
/product="hypothetical protein, weakly similar to
(AF102871) neuronal apoptosis inhibitory protein 2 [Mus
musculus]"
/protein_id="CAB97523.1"
/db_xref="gi:9367840"
/translation="INSGNITPDYLFDFEHLPCNSALDITKIDFYGGAMASWEKAAE
DTGILHMEAEPTYIPSAVSLFPMWKEPRTLEVTLDTSKLNKODIRYLGKLFSSA
TSLRQIKRCAGVYSLSVLSTCKNITSLSVVEASPLTIEDERHTSVNKTLSIH
LONORLPGILDSGLNKLTKLIMDNIMNEEDAIKAEGLKLNKCPHLHLISD
IGEGMDYIVKSSSEPCDLERLQVSCLSANAVKIIAONLHNILVKSILDSSEYLE
KQGNALHEILIDRMNVLEOLITAIMPMGCDVGSISLTKLELVEPOLVYKGIKRMRL
TDTETRIIGAFEPKPKLPKNFOOLNLAGNRVSSDGLAWGTFENIKQIVLFDFTKEE
LPDPALVKKLSQVSKLFTQEARLVGWQFDDDLSTVITGAFKLVTAA"

CDS
BASE COUNT 416 a 265 c 312 g 362 t
ORIGIN
alignment_scores: Quality: 2244.00 Length: 439
Ratio: 5.123 Gaps: 0
Percent Similarity: 99.772 Percent Identity: 99.772

alignment block:
US-09-697-089-2 x IR2005417 ..
Align seg 1/1 to: IR2005417 from: 1 to: 1355

586 ILeasSercIyAsnIleProaSerTyrlEuPheaspPheaglIuhIste 602
|||||
2 ATCAACTCAGGGAACATCCCGATTACTTATTGACTCTTTGAACATTT 51
|||||
602 uProaScyAlaSerAlaLeuaspPheIlelySleuaspPheTygIyG 619
|||||
52 GCCCAATTGGCAAGTGCCTGCACTTCATTAACTGACCTTTATGGGG 101
|||||
619 lYalameAlaSerTPpGluYAlaAlaGluaspThrglyIleHs 635
|||||
102 GACCTATGGCTTCATGGAAAAGCGTCAGAAACACAGAGTGGAAATCCAC 151
|||||
636 MeGluGluAlaProgluThrTyrlProserArAlaValSerleuH 652
|||||
152 ATGGAAGAGGCCCGAAGACCTACATTCACGACGAGGCTGTATCTTGT 201
|||||
652 ePheasTrpLySGluGluPheArgrThrleuGluValThrleuArGasp 669
|||||
202 CTTCACCTGGAGCAGGAATTGAGACTCTGGAGTCCACACTCCGGATT 251
|||||
669 heserIySleuasnlySGlnaspIleThrTyrlleuGlylySlePheSer 685
|||||
252 TCAGCAAGTTGAATTAAGCAAGATATCATCTGGGAAAATATTCACG 301
|||||
686 SerAlaThrSerleuArgrleuGluInilelyArGyAlaGlyValAlaGl 702
|||||
302 TCTGCCACAAAGCTCAGGCTGCATAATAAGATGTGCTGCTGTGCTGG 351
|||||
702 ySerleuSerleuValleuSerThrCyslySasnIleTySerleuMetY 719
|||||
352 AACCTCAGTTGTGCTCAGACACTGTAAAGAACATTTATCTCATCATG 401
|||||
719 alGluAlaSerProleuThrIleGluaspGluArghisIleThrSerVal 735
|||||
402 TGGAAACCAAGTCCCTCACCATTAAGAGATGAGGACATCATCTGTA 451
|||||
736 ThrAsnleuThrleuSerIleHsAspIleuGlnAsnGluArgrleuP 752
|||||
452 ACAAACTGAAAACCTTGAGTATTCTATGACTTACAAATCAACAGGCTGCC 501
|||||
752 ogIglyleuThrAspSerleuGlyAsnleuLySasnleuThrlySleuI 769
|||||
502 GGGTGTCTGACTGACAGCTGGGTAACTTGAAGAACCTTACAAAGCTCA 551
|||||
769 lMetaspasnIlelyMetasnGluaspAlaIlelySleuAlaGlu 785
|||||
552 TAATGATTAACATTAAGATGAAGAAAGATGCTATAAACTAGCTGAA 601
|||||
786 GlYleuLySasnleuLySlyMetCysleuPheHsIleuThrHsIleuSe 802
|||||
602 GGCCTAAAAAAGCTGAAGAAGATGCTTTTATTTCAATTTGACCCACTTGTG 651
|||||
802 rasPIleglyGluGlyMetaspTyrlIleVallySerleuSerSerGluP 819
|||||
652 TGACATTTGAGAGGAGGAATGATATACATAGCAAGCTCTGTGTAAGTGAAC 701
|||||
819 rOcysaspIleuGluGluIleGlnleuValSerCysCysleuSerAlaasn 835
|||||
702 CCGTGATACCTTGAAGAAATTAATTAAGTCTCCGCTGCTGTGCAAAAT 751
|||||
836 AlaVallySIlleuAlaGlnAsnleuHsAsnleuVallySleuSerIl 852
|||||
752 GCAGTGAATAATCTAGCTCAGAACTTTCACATTTTGGTAAACTGAGCAT 801
|||||
802 TCTTGAATTTATGAGAAATTAACCTGGAAGATGAATGAAGCTCTTC 851
|||||
869 isgluIleuIleaspArgMetasnValleuGluGlnleuThralaIleuMet 885
|||||
852 ATGAACTGATCGACAGATGAAGCTGTAGAACAGCTCACCCATGAGAG 901
|||||
886 leuProTrpGlyCysaspValGlnGlySerleuSerSerleuIleuIlyshI 902

```

1002 CCGCCCTGGGGCTGTGAGCTTCAGGCGAGCGCTGAGCGCTGTTGAAACA 951
1002 sLeuGlulValProGlnLeuValLysLeuGlyLeuLysAsnTrpArg 919
902 sLeuGlulValProGlnLeuValLysLeuGlyLeuLysAsnTrpArg 919
952 TTTGGAGAGGCTCCACAACTGTCAGAGCTGGGTTGAAAACTGGAGAC 1001
919 eutHrAspThrGlnIleArgIleLeuGlyAlaPhePheGlyLysAsnPro 935
1002 TCACGATACAGAGATTAGAAATTTTACGTCATTTTGGAAAGAACCCCT 1051
936 LeuLysAsnPheGlnGlnLeuAsnLeuAlaGlyAsnArgValSerSerAs 952
1052 CCGAAAAAATTCACAGCACTTCATATTTGGCGGAAATGCTGTGACCACTGA 1101
952 pGlyTrpLeuAlaPheMetGlyValPheGluAsnLeuLysGlnLeuValP 969
1102 TCGATGCGCTGCTCATGCGTGATTTGAGAAATCTTAACCAATTTAGTGT 1151
969 hePheAspPheSerThrIrysgLuePheLeuProAspProAlaLeuValArg 985
1152 TTTTTCGACTTTAGTACTAAGAAATTTCTACCTGTCACAGATTAGCAGA 1201
986 LysLeuSerGlnValLeuSerLysLeuThrPheLeuGlnGlnAlaArgLe 1002
1202 AAACCTTACCACCAAGTGTATCCCAAGTTAACTTTTTCGACAAACACTAGGCT 1251
1002 uValGlyTrpGlnPheAspAspAspLeuSerValIleThrGlyAlaP 1019
1252 TGTTCGGTGGCAATTGTGATGATGATGATCAGTGTTATACAGGTGCTT 1301
1019 heLysLeuValThrAla 1024
1302 TTAACCTAGTAACTGCT 1318

seq_name: gb_pr:AY027788

seq.documentatation_block: 1395 bp mRNA PRI 20-JUL-2001
LOCUS AY027788 Homo sapiens CLANB (CLAN1) mRNA, complete cds.
DEFINITION Homo sapiens CLANB (CLAN1) mRNA, complete cds.
ACCESSION AY027788
VERSION AY027788.1 GI:14324114
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1395)
AUTHORS Damiano,J.S., Stehlik,C., Pio,F., Godzik,A. and Reed,J.C.
TITLE Clan, a novel human ced-4-like gene
JOURNAL Genomics, 75 (1-3), 77-83 (2001)
MEDLINE 21365712
PUBMED 11472070
2 (bases 1 to 1395)
AUTHORS Stehlik,C., Damiano,J.S., Pio,F., Godzik,A. and Reed,J.C.
TITLE Direct Submission
JOURNAL Submitted (21-FEB-2001) Program on Apoptosis and Cell Death
Research, The Burnham Institute, 10901 North Torrey Pines Road, La
Jolla, CA 92037, USA
FEATURES
source
1. .1395
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2p22-p21"
/tissue_type="lung"
1. .1395
/gene="CLAN1"
/gene=".1356
277. .1356
/codon_start=-1
/product="CLANB"
/protein_id="AAK14777.1"

```

```

/db, xref="GI:14324115"
/translation="MNRITKDNRSALITRMGNTYIKQTTDLFVWNVUNREEVNITCE
KVEDAARGLIIMILKRGSSCNLFELSLKEWNPPLFQDINGSGSLTDSIGNLKITK
LIMONIKANWEDAIKLEGLKLNKMKCLFHLTHLSDEGMQYIVKLSSEPCDLEI
QVSCCLSANAVKILAONLHNLYKLSITDSENVLEDMGDEALHLEIDRNNVLEQLTA
LMPWGCDVOGSLSLTKLHEEYVQPLKFGIKMNRLLDTERTIGAFGKKNPLKPOQ
LNLAGNVSSDGMFLAEFGVFNPKQVLFPPFPERKEFLPDPAIVLRKLSQVSLKTLFQE
ARLVGMPFDDDDLSVITGAKRLVTA"
BASE COUNT      436 a      248 c      327 g      384 t
ORIGIN

alignment_scores:
      Quality: 1500.50      Length: 1024
      Ratio: 4.180      Gaps: 1
Percent Similarity: 35.059      Percent Identity: 35.059

alignment block:
US-09-697-089-2 x AY027788      ..

Align seg 1/1 to: AY027788 from: 1 to: 1395

1 MetAsnPheIleLysAspAsnSerArgAlaLeuIleGlnArgMetGlyLe 17
|||||
277 ATGATTTTATTAAGGACATATGACCGACCCCTTATTCAAGATGGAGAT 326

17 LThrValIleLysGlnIleThrAspAspLeuPheValTrpAsnValLeu 34
327 GACTGTATTAAGCAATCATCAGATGACCTATTGTGATGGATGTTCTGA 376

34 snArgGluGluValAsnIleIleCysCysGluLysValGluGlnAspAla 50
|||||
377 ATCGCCAAAGAGTAAACATCATTTGCTGCGAAGGTGGAGCAGAGTGCT 426

51 AlaArgGlyIleIleHisMetIleLeuLysGlySerGluSerCysAsp 67
|||||
427 GCTAGAGGATCATTCACATGATTTTGAAATAAGGGTTCCAGAGTCCGTGA 476

67 nLeuPheLeuLysSerLeuLysGluTrpAsnTrpProLeuPheGlnAspL 84
477 CCGTCTTTTAATCCCTTAAGAGAGTGAACATCCGTATTTCAAGACT 526

84 euAsnGlyLysSerLeuPheHisGlnThrSerGluGlyAspLeuAspAsp 100
|||||
527 TGAATGACACAAGT..... 540

101 LeuAlaGlnAspLeuLysAspLeuTrpHisThrProSerPheLeuAsnPh 117
540 ..... 540

117 eTrpProLeuGlyGluAspIleAspIleIlePheAsnLeuLysSerThrP 134
540 ..... 540

134 heThrGluProValLeuTrpArgLysAspGlnHisHisArgValGlu 150
540 ..... 540

151 GlnLeuThrLeuAsnGlyLeuGlnAlaLeuGlnSerProCysIleIle 167
540 ..... 540

167 eGluGlyGluSerGlyLysGlyLysSerThrLeuLeuGlnArgIleAlaM 184
540 ..... 540

184 eLeuTrpClySerGlyLysCysLysAlaLeuThrLysPheLysPheVal 200
540 ..... 540

201 phePheLeuArgLeuSerArgAlaGlnGlyGlyLeuPheGluThrLeuCy 217
540 ..... 540

```


217 saspGlnLeuLeuaspIleProGlyThrIleArgLysGlnThrPheMet 234
 540 540
 234 IamLeuLeuLysLeuArgGlnArgValLeuPheLeuLeuaspGlyTyr 250
 540 540
 251 AsnGlnPheLysProGlnAsnCysProGluIleGluAlaLeuIleLysGlu 267
 540 540
 267 uasnHisArgPheLysAsnMetValIleValThrThrThrGlnCysL 284
 540 540
 284 euArgHisIleArgGlnPheGluAlaLeuThrAlaGluValGlyAspMet 300
 540 540
 301 ThrGluAspSerAlaGlnAlaLeuIleArgGluValLeuIleLysGluLe 317
 540 540
 317 uAlaGluGlyLeuLeuLeuGlnIleGlnLysSerArgCysLeuArgAsnL 334
 540 540
 334 euMetLysThrProLeuPheValAlaIleThrCysAlaIleGlnMetGly 350
 540 540
 351 GluSerGlnPheHisSerHisThrGlnThrThrLeuPheHisThrPheTyr 367
 540 540
 367 rAspLeuLeuIleGlnLysAsnLysHisLysHisLysGlyValAlaAlaAs 384
 540 540
 384 eraSpheIleArgSerLeuAspHisCysGlyAspLeuAlaLeuGluGly 400
 540 540
 401 ValPheSerHisLysPheAspPheGlnLeuGlnAspValSerSerValAs 417
 540 540
 417 nGluAspValLeuLeuThrThrGlyLeuLeuCysLysTyrThrAlaGlnA 434
 540 540
 434 rgPheLysProLysTyrLysPhePheHisLysSerPheGlnGluTyrThr 450
 540 540
 451 AlaGlyArgArgLeuSerSerLeuLeuThrSerHisGluProGluGluVal 467
 540 540
 467 lThrLysGlyAsnGlyTyrLeuGlnLysMetValSerIleSeraspIleT 484
 540 540
 484 hrSerThrTyrSerSerLeuLeuArgTyrThrCysGlySerSerValGlu 500
 540 540
 501 AlaThrArgAlaValMetLysHisLeuAlaAlaValTyrGlnHisGlyCys 517
 540 540

517 sLeuLeuGlyLeuSerIleAlaLysArgProLeuTyrPArgGlnGluSerL 534
 540 540
 534 euGlnSerValLysAsnThrThrGlnGlnGlnIleLeuLysAlaIleAsn 550
 540 540
 551 IleAsnSerPheValGluCysGlyIleHisLeuTyrGlnGlnSerThrSe 567
 540 540
 567 rLysSerAlaLeuSerGlnGlnPheGluAlaPhePheGlnClyLysSerL 584
 540 540
 584 euTyrIleAsnSerGlyAsnIleProaspTyrLeuPheaspPheGln 600
 540 540
 601 HisLeuProaspCysAlaSerAlaLeuAspPheIleLysLeuaspPheTyr 617
 540 540
 617 rGlyGlyAlaMetAlaSerTrpGluLysAlaAlaGluAspThrClyGlyI 634
 540 540
 634 leHisMetGluGluAlaProGluThrTyrIleProSerArgAlaValSer 650
 540 540
 651 LeuPhePheAsnTrpLysGlnGluPheArgThrLeuGluValThrLeuAr 667
 540 540
 667 gAspPheSerLysLeuAsnLysGlnAspIleThrTyrLeuGlyLysIleP 684
 540 540
 684 heSerSerAlaThrSerLeuArgLeuGlnIleLysArgCysAlaGlyVal 700
 540 540
 701 AlaGlySerLeuSerLeuValLeuSerThrCysLysAsnIleTyrSerLe 717
 540 540
 717 uMetValGluAlaSerProLeuThrIleGluAspGluArgHisIleThrS 734
 540 540
 734 erValThrAsnLeuLysThrLeuSerIleHisAspLeuGlnAsnGlnArg 750
 540 540
 751 LeuProGlyGlyLeuThrAspSerLeuGlyAsnLeuLysAsnLeuThrLy 767
 541 541
 767 sLeuIleMetAspAsnIleLysMetAsnGluGluAspAlaIleLysLeuA 784
 582 GCTCATTAATGATACATTAAAGATGAATGAAGAAGATGCTATTAAGACTAG 631
 632 CTGAAGGCTGAAACCTGAAGAAGATGCTTATTTATTGATTGACCCAC 661
 801 LeuSeraspIleGlyGluGluMetAspTyrIleValLysSerLeuSerSe 817
 682 TTGCTGTACATTGGAGGGAATGGAATTACATGATCAAGTCTGTCTGCAAG 731
 817 rGluProCysAspLeuGlnGluIleGlnLeuValSerCysCysLeuSerA 834

```

|||||
732 TCAACCCCTGACCTTGAGAAATTCATTAGCTCTCTCTCTGCTG 781
834 TAAATGATGATGATGATGATGATGATGATGATGATGATGATG 850
782 CAAATGATGATGATGATGATGATGATGATGATGATGATGATG 831
851 SerileuAspLeuSerGluAsnTyrLeuGluTyrAspGlyAsnGluAl 867
832 ACCATTCTTGATTTATCAGAAAATTACCTGGAAGAAAGATGGAATAC 881
867 AleuHisGluLeuIleAspArgMetAsnValLeuGluGlnLeuThrAla 884
882 TCTTCATGAACTGATCGACAGATGAACTGCTAGAACAGCTCACCGCAC 931
884 eumelleuprotirpGlycysaspValGlnGlySerLeuSerSerLeu 900
932 TATCTGCTGCTGGGGCTGTGACGTCAAGCAAGCAAGCTGAGAGCTGTG 981
901 LysHisLeuGluGluValProGlnLeuValLysLeuGlyLeuLysAsn 917
982 AAACATTTGGAGAGAGTCCACACACTCGTCAAGCTTGGGTGAAAACG 1031
917 PArgLeuThrAspThrGluIleArgIleLeuGluAlaPhePheGlyLys 934
1032 GAGACTCACAGATACAGAGATGAAATTTAGTGCTATTTTGGAAAGA 1081
934 snProLeuLysAsnPhenGlnGlnLeuAsnLeuAlaGlyAsnArgVal 950
1082 ACCCTCTGAAAACCTTCCAGCACTTGAAATTTGGCGGAAATGCTGTG 1131
951 SeraspGlyTyrPheValaPheMetGlyValPheGluAsnLeuLysGln 967
1132 ACTGATGATGATGCTTCCCTTCAATGGCTGTATTGAGATCTTAAGCA 1181
967 ValaPhePheAspPheSerThrLysGluPheLeuProAspProAlaLeu 984
1182 AGGTGTTTTGACTTTAGTAAAGAAATTTCTACCTGATCCAGCACTG 1231
984 AlArgLysLeuSerGlnValLeuSerLysLeuThrPheLeuGlnGluAl 1000
1232 TCAGAAAACCTTACCCAGTGTATCCAACTTAACCTTTCTGCAGAGCT 1281
1001 ArgLeuValGlyTyrPheGlnPheAspAspAspLeuSerValIleThr 1017
1282 AGCTTTGTTGGGTGGCAATTTGATGATGATGATGATGATGATGATG 1331
1017 ValaPheLysLeuValThrAla 1024
1332 TGCCTTTAAACTAGTAACTGCT 1353
-g-name: gb_pr:AY027789
In documentation block:
FEATURES             768 bp      mRNA      PRI      20-JUL-2001
DEFINITION Homo sapiens CLANC (CLAN1) mRNA, complete cds.
ACCESSION AY027789
VERSION AY027789.1 GI:14324116
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Damiano,J.S., Stehlik,C., Plo,F., Godzik,A. and Reed,J.C.
TITLE Clan, a novel human ced-4-like gene
JOURNAL Genomics. 75 (1-3), 77-83 (2001)
MEDLINE 21365712
PUBMED 11472070
REFERENCE 2 (bases 1 to 768)
AUTHORS Stehlik,C., Damiano,J.S., Plo,F., Godzik,A. and Reed,J.C.
TITLE Direct Submission
JOURNAL Submitted (21-FEB-2001) Program on Apoptosis and Cell Death

```

```

FEATURES
SOURCE Research, The Burnham Institute, 10901 North Torrey Pines Road, La
        Jolla, CA 92037, USA
        Location/Qualifiers
            1..768
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="2"
            /map="2p22-p21"
            /tissue-type="lung"
            1..768
            /gene="CLAN1"
            277..747
            /gene="CLAN1"
            /codon_start=1
            /product="CLANC"
            /protein_id="P4K14778.1"
            /db_xref="GI:14324117"
            /translation="MNFKDNRRALLQRMGMVYIKQITDDLFWYVNLREEVNIICE
            KVEDDAARGIHHMLKKSSCNLFSLKEMVPLFODLNGSLFPHOTSGDLDDIA
            ODLKDLTFPSFLNFPYPLGSDIDIFPKSTFEPVLMRKDOHHHREQLTVL"
BASE COUNT 218 a 157 c 180 g 213 t
ORIGIN
alignment_scores:
    Quality: 819.00      Length: 154
    Ratio: 5.318        Gaps: 0
Percent Similarity: 100.000    Percent Identity: 100.000
alignment_block:
US-09-697-089-2 x AY027789
Align seg 1/1 to: AY027789 from: 1 to: 768
1 MetAsnPhelIleLysAspAsnSerArgAlaLeuIleGlnArgMetGlyLe 17
277 ATGAATTTTATTAAGACATAGACCGAGCCCTTATTCAAGATGGGAAT 326
17 ThrValIleLysGlnIleThrAspAspLeuPheValTyrPheValLeu 34
327 GACTGTTATAAGCAATCAGATGACCTATTGTATGATGATGATGATG 376
34 snArgGluGluValaValaValIleLysCysGluLysValGluVala 50
377 ATCGCGAAGAGTAAACATCATTTGCTCGAGAGAGTGGAGCAGATGCT 426
51 AlaArgGlyIleIleHisMetIleLeuLysLysGlySerGlySerCys 67
427 GCTAGAGGATCATTCACATGATTTTGAAGAAAGGTTTCAGAGTCTGT 476
67 nLeuPheLysSerLeuLysGluTyrPheAsnTyrPheLeuPheGlnAsp 84
477 CCTCTTCTTAATCCCTTAAGAGTGAAGTATCTCTATTTCAAGACT 526
84 euAsnGlyLysSerLeuPheHisGlnThrSerGluLysAspAspAsp 100
527 TGAATGACAAAGTCTTTTTCATCAGACATCAGAAAGAGACTTGGAC 576
101 LeuAlaGlnAspLeuLysAspLeuTyrHisThrProSerPheLeuAsn 117
577 TTGGCTCAGAGATTTAAAGACTTGTACATACCCATCTTTCTGAACT 626
117 eTyrProLeuGlyGluAspIleAspIlePhePheAsnLeuLysSerThr 134
627 TTAATCCCTTGGTGAAGATATGACATTTTATTAACTGAAAGACACT 676
134 heThrGluProValLeuTyrPheLysAspGlnHisHisValGlu 150
677 TCACAGAACCTGCTCGTGGAGGAAGACCAACACATCATCCGCGAG 726
151 GlnLeuThrLeu 154
727 CAGCTGACCCCTA 738

```

seq_name: gb_pat:A64531

seq_documentation_block: 6228 bp DNA PAT 29-MAR-1999

LOCUS A64531 Sequence 23 from Patent WO9726331.

DEFINITION A64531

ACCESSION A64531

VERSION A64531.1 GI:3717929

KEYWORDS

SOURCE unidentified.

ORGANISM unidentified.

REFERENCE 1 (bases 1 to 6228)

AUTHORS Korneluk, R.G., Mackenzie, A.E., Roy, N., Robertson, G. and Tama, K.

TITLE USE OF NEURONAL APOPTOSIS INHIBITOR PROTEIN (NALP)

JOURNAL Patent: WO 9726331-A 23 24-JUL-1997;

COMMENT UNIV OTTAWA (CA)

FEATURES

Other publication AU 1614997 19970811.

source 1..6228 Location/Qualifiers

BASE COUNT 1825 a 1332 c 1307 g 1764 t

ORIGIN

alignment_scores:

Quality: 655.00 Length: 1108

Ratio: 1.077 Gaps: 42

Percent Similarity: 54.874 Percent Identity: 23.736

alignment_block:

US-09-697-089-2 x A64531 ..

Align seq 1/1 to: A64531 from: 1 to: 6228

69 PheLeuysSerLeuLysGluThrPAsnTyrProLeupheGlnAspLeuAs 85

1425 TTTTCCAAATATAGAACTCC...TTCGCGAAGTGAATCCAGACCTTCA 1471

85 nglYgln.....SerLeupheHsGlnThrSerGluGlyAspL 98

1472 GAGCGGTGGTGAAGTCTTGTAATTAAGTGAAGCAAGTGAAGCAATC 1521

98 euAspAsp..... 100

1522 TTGAAGATTCAATAGCAGTTGCTCTAAGTCCAGAAATGCGACAGGTT 1571

101LeuAlaGlnAspLeuLysAs 107

1572 GAACCCAGTGTTCAGAGAGCAAGAAATCTGAATGACGACGCTGACAGC 1621

107 PLeuTyrHisThrProSerPheLeuAsnPhe..... 117

1622 AGCTTATACAGCGCCAGTTTCCGCAATGCTTTGCTGATATCTCTT 1671

118TyrProLeuGlnGluAspLeuAspLeuLeuPhe 128

1672 CCGATTCGGCCAGCGACACTGCTGGGCTGTGATCTGTAT..... 1715

129 AsnLeuLysSerThrPheThrGluProValLeuThrPArgLysAspGlnH 145

1716 ...GCTTCAAACACATCAGCAAACTGTG..... 1742

145 SHSHSHArGValGluGlnLeuThrLeuAsnGlyLeuLeuGlnAlaLeuG 162

1743CAGAACCTCTGCTGTCTGCTGAGTCTTTGGCAACTTGA 1782

162 InseProGlyIleIleGluGlnGlySerGlyLysSerThrLeu 178

1783 ACTGTGCATGTGTGTGAGAGGAGTGAAGTGAAGAGAGGCTCTC 1832

179 LeuGlnArgIleAlaMetLeuThrGlySerGlyLysAlaLeuThr 195

1833 CTGAGAAATAGCTTTCTGTGGCATCTGGATGCTGCCCTGTAA 1882

195 rLysPheLysPheValPhePheLeuArgLeuSer.....ArgAlaGln 210

1883 CAGGCTTCAGCTGTGTCTTACCTCTCCCTTAGTTCACACAGACAGC 1932

210 LysGlyLeuPheGluThrLeuLysAspGlnLeuLeuAspIleProGlyThr 226

1933 AGGGCTGGCCAGTATATCTGTGACAGCTCTCAGAGAAAGAGATCT 1982

227 IleArgLysGlnThrPheMetAlaMetLeuLeuLysLeuArgLysArgVal 243

1983 GTTACGTGAATGTGCATGAGAAACATTATCCAGAGTTAAAGATCAGT 2032

243 LeuPheLeuLeuAspGlyTyrAsnGluPhe.....LysProGlnAsn 258

2033 CTATATCTCTTTAGATGATCAAGAAATATGTTCAATCCCTCAA.... 2078

258 ysProGluIleGluAlaLeuIleLysGluAsnHisArgPheLysAsnMet 274

2079GTCAATAGGAATCTGATTCAAAATAACCTTATCCCGACCTGC 2123

275 ValIleValThrThrThrThrGluLysLeuArgHisIleArgGlnPheG 291

2124 CTATGATTCGCTGCTCCGTACAAACAGGCGCAGCATCCGCGATACCT 2173

291 yAlaLeuThrAlaGluValGlyAspMetThrGluAspSerAlaGlnAla 308

2174 AGAGACCATTTCTAGATCCAGCATTTCCCTTTATATATCTGTCTGA 2223

308 euIleArgGluValLeuIleLysGluLeuAla.....GluGlyLeuLeu 322

2224 TATTACGGAAGCTCTTTCAATATATATGACTGCTGCGAAAGTTATG 2273

323 LeuGlnIleGlnLysSerArgCysLeuArgAsnLeuMetLysThrProle 339

2274 GTTACTTGTGGAAGAACCAAGTTTGCAGAGATATAGAAATCTCTCT 2233

339 upheValValIleThrCysAlaIleGlnMetGlyLysGluPheHis 356

2324 CTGTGTGGCGCGATCTGTGCTCATTTGCTTCACTATCTTTGACCCAT 2373

356 eHisThrGlnThrThrLeuPheHisThrPheThrPheLysPheLeuIleGln 372

2374 CTTTGTGATGCTGCTGTTTTCAGTCTTATATGAGACGCTTCTCTTA 2423

373 LysAsnLysHisLysHisLysGlyValAlaAlaSerAspPheIleArg.. 388

2424 AGGACAAA.....GCGACACTGCAAAATTCGCAAGC 2455

389 SerLeuAspHisCysGlyAspLeuAlaLeuGluGlyValPheSerHisL 405

2456 AACGTGTCTCTCTGCTGTGAGCTGCGCTTGAAGGCTTTTTCATGTT 2505

405 ysPheAspPheGluLeuGlnAspVal.....SerSerValAsnGluAsp 419

2506 GCTTTGAGTTTATGATGATGATCTCCAGACAGGAGGTTGATGAAGAT 2555

420 ValLeuLeuThrThrGlyLeuLysLysLysTyrThrIleArgIleArgPhe 436

2556 GAAGATCTAACCATGTCTGATGAGCAAAATTTACACCCAGAGACTAAG 2605

436 sProLysTyrLysPhePheHisLysSerPheGlnGluTyrThrAlaGly 453

2606 ACCATTCCTACCGGTTTTTAAGTCTCTCCAGAAATTTCTTGGGGGA 2655

453 rGArgLysSerSerLeuLeuThrSerHisGluProGluGluValThrLys 469

2656 TGAAGCTGATTAAGATCTCTGATTCAGATAGCGAGAACATCAAGATTG 2705

470 GlyAsnGlyTyrLeuGlnLysMetValSerIleSerAspIleThrSerTh 486

BASE COUNT	1777 a	1267 c	1267 g	1673 t
ORIGIN				

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

275 ValIleValThrThrGluCysLeuArgHisIleArgGlnPheG1 291

```

2020 CTATTGATGTCGTCGTAACAACAGGGCCAGGACATCCGCCGATACCT 2069
2021 YAlaLeuThrAlaGluValAlaGlyAspMetThrGluAspSerAlaGln 308
2070 AGAGACCATTCAGAGATCAAGACATTTCCCTTATATATACTGCTGTGA 2119
308 euLeaArgGluValLeuLeuGluLeuVala.....GluLeuLeu 322
2120 TATTACGGACCTCTTTACATATATATGACCTCGCTCGAAGATTTATG 2169
323 LeuGlnIleGlnLysSerArgCysLeuArgAsnLeuMetLysThrProle 339
2170 GTTACTTGGAAAGACCAAGATTGGACAGATACAGAAACCTCCT 2219
339 uPheValValIleThrCysAlaIleGlnMetGlyLysSerGluPheHis 356
2220 CTTTGTGGCGGCGATCTGTGTCATGTTTTCAGTATCCTTTTGACCAT 2269
356 eRhIstHcIleThrThrLeuPheHisThrPheTyrAspLeuLeuIleGln 372
2270 CCTTGTGATGATGGCTGTTTTCAGTCTTATATGAAACGCTTTCCCTTA 2319
373 LysAsnLysHisLysHisLysGlyValAlaAlaSerAspPheIleArg.. 388
2320 AGGACCAAA.....GCGACAGCTGAATTCGAAAGC 2351
389 .SerLeuAspHisCysGlyAspLeuAlaLeuGlnIleValPheSerHisL 405
2352 AACGTGTCCTCCTGTGGTGAAGCTGGCTTGAAAGGCTTTTTCATGTT 2401
405 yspHAspPheGluLeuGlnAspVal.....SerSerValAsnGluAsp 419
2402 GCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2451
420 ValLeuLeuThrThrGlyLeuLeuCysLysTyrThrAlaGlnArgPhe 436
2452 GAAGATCTTAACCATGCTGCTGATGAGCAATTTACAGCCAGACTAAG 2501
436 sProLysTyrLysPhePheHisLysSerPheGlnIleTyrThrAlaGly 453
2502 ACCATTCTACCGCTTTTAAGTCCTGCCCTCCAGAAATTTCTGGGGGA 2551
453 rGArgLeuSerSerLeuLeuThrSerHisGluProGlnGluValThrLys 469
2552 TGAGGCTATGAACTCTGATTCAGATAGCGAAGACATCAAGATTG 2601
470 GlyAsnGlyTyrLeuGlnLysMetValSerIleSerAspIleThrSerTh 486
2602 GGACTGATCATTTGAACAATCACTCAACCATGATGACTGTAAAGCGC 2651
486 rTyrSerSerLeuLeuArgTyrThrCysGlySerSerValGluAlaThr 503
2652 CTACAAACAATTTTGAACATATGC.....TCCAGCTCCCTTCACAA 2695
503 rGAla.....ValMetLysHisLeuAlaValTyrGlnHisGly 516
2696 AAGCAGGCGCCAAATTTGTCCTCATTTGCTCCATTAGTGGATTAAC... 2742
517 CysLeuLeuGlyLeuSerIleAlaLysArgProLeuTyrArgGlnLys 533
2743 .....AAAGAGTC 2750
533 rLeuGlnSerValLysAsn..... 539
2751 ATTGGAGATATATCTGAATAATGATGACTACTTAAGCAGCAGCAAAA 2800
540 ..ThrThrGluGlnGluLeuLeuLysAlaIle.....AsnIle 551
2801 TTTCACTCGCATGCGATTACTTAGGGATTTGCAAAATTTGTCCACAA 2850
552 AsnSerPheValGluCysGlyIleHisLeu..... 561
2851 GCTTACTTTTCAATGGTTTCAGAACATTTACTGTCTTCCCTGAAAC 2900
562 ...TyrGlnLysSerThrSerLysSerAlaLeuSerGlnIlePheGlu 577
2901 TGCTTATCAAGACACT...GTTCGTCGGTGTCTTCATTTGTTTGC 2947
577 lAphePheGlnGlyLysSerLeuTyrIleAsnSerGlyAsnIleProAsp 593
2948 AATTCCTTCAAGGAGAACACTGACTTTGGGTGCCTTAACCTTA...CAG 2994
594 TyrLeuPheAsp..... 597
2995 TACTTTTGGACCCACAGAAAGCTTTCATGTTTGGAGACATCCACTT 3044
598 .....PhePheGlnHisLeuP 603
3045 CCCATAGCAGGAATAAGACATCCACAGACACATTTTTCAGTTCG 3094
603 rOAsnCys.....AlaSerAlaLeuAspPheIleLysLeuAspPheTyr 617
3095 AAACATGTTTGGACAAATACACAGGTGCCAATATGATCAGGACTATGCT 3144
618 GlyGly.....AlaMetLaserThrPglu..... 625
3145 TCTGCTTTGAACCTATATGATGATGGAGCGAAATTTAGCTGAAAAAGA 3194
626 .....LysAlaAlaGluAspT 631
3195 GGATATATGTAAGACCTATATGATATGACAGCGACGATCCACAGAC 3244
631 hrGlyGlyIleHisMetGluGluAlaProGluThrTyr...IleProSer 646
3245 TTAGTACGCTGATTTGAAACCTTCTCCAAAGACGTACAAATGCC... 3291
647 ArgAlaValSerLeuPhePheAsnThrLysGlnIlePheArgThrLeuG 663
3292 .....TGTCTAGA 3299
663 uValThrLeuArgAspPheSerLysLeuAsnLysGlnAspIleThrTyrL 680
3300 AGTCATGTCAGATGATATGATGATGATGATGATGATGATGATGATGAT 3349
680 euGlyLysIlePheSerSerAlaThrSerLeuArgLeuGlnIleLysArg 696
3350 TAAATGACAGTTTCTACACTTCACAGCGCATCGAATCTCAATTAACAC 3399
697 CysAlaGlyValAlaGlySerLeuSerLeuValLeuSerThrCysLys.. 712
3400 AGCAGAGGCTTTATGAAGAATCGCCAGCTCTGTAGCTGTCTAAGGC 3449
713 .AsnIleTyrSerLeuMetValGluAlaSerProLeuThrIleGluAsp 729
3450 CTCTGTACCAAGATGCTCATAGCAAGTTGGAATCTGACGCGACGCAAC 3499
729 lUAArgHisIleThrSerValThrAsnLeuLysThrLeuSerIleHis... 744
3500 AGGAACCTCTTCTCACCTGCTTCCCTCGAATCTCTGAAGTCTCAGCG 3549
745 .....AspLeuGlnAsnGlnArgLeuProGlyGlyLeuThrAspSerLe 759
3550 ACAATCCAGTCAACAGACCAATCTTCTCTAT.....CTGGATTAAGTT 3593
759 uGlyAsnLeuLysAsnLeuThrLysLeuIleMetAspAsnIleLysMet. 775
3594 CCTGTGCTGAAGACAGTCTGTGATCTGTGAGGCGCAATTAATGTTT 3643
775 ..... 775
3644 TTTCACTATTCCTGAAAGATTTCCAAACTTCACACATATGAGAAATTA 3693
776 .....AsnGluGluAspAlaIleLysLeuAlaGluGlyLe 787
3694 TTTGATCAAAATTTCAAGCTGAGATGATCTTCCAAACTAGTAATTAAT 3743

```



```

787 ulysasnleuylslysmetCysleupheHisLeu.....ThrHisLeuS 802
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
3744 TCMAAAATTCACCAACCTTCATGTTTCCATCTGAACTGTAACCTCTTT 3793
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
802 eraSpIleGIugIugIyMetAspTyrIleValIysSerLeuSerSerGIu 818
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
3794 CGGATTTTGGGTCTCATGACTATGCTGTTC..... 3828
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
819 ProCys...AspleuGIuGIleGIleuValSerCysCysLeuSerAl 834
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
3829 ...TGTAAAGAACTCACAGAAATTAAGTTTCGATTCATTTTT..... 3870
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
834 aasnAlaValIysIleLeuAlaGlnAsnLeuHisAsnLeuValIysLeuS 851
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
3871 .CAAGCCGTCCTCCATTGTTGCC...AGTTGCCAAATTTATTCTCTGTA 3916
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
851 erIleLeuAspLeu...SerGIuAsnTyrLeuGIuIysAspGIyAsnGIu 866
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
3917 AGATATTAAATCTTGAAGCCAGCAATTCCTGTGATGAGAAACATCGAA 3966
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
867 AlaLeuHisGIuLeuIleAspArgMetAsnValLeuGIuGIuLeuThrAl 883
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
3967 AAATTTGCCCTACATTTTAGCTTCTTAGTACCTGAGAA..... 4008
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
883 alaumeLeuProTropGIyCysAspValGIuGIySerLeuSerSerLeuL 900
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
4009 .TTGATCTCTTCTACTGCGGATGGAATTTTCAGTGCGCCAAACGTATCA 4057
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
900 eulysHisLeuGIuGIuValProGIuLeuValIysLeuGIyLeuLysAsn 916
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
4058 TCCAGCAGTGTACAGACCTTCATGTCCTCGAGCTCTCATTTTTCAG 4107
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
917 TrpArgLeuThrAspThr.....GluIleArgIleLeuGI 928
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
4108 ACCTTGAATGATGACAGCGGTGTGAAATTCACAAAGTACAAATCAAGTG 4157
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
928 yAlaPhePheGIyLysAsnProLeuLysAsnPheGIuGIuLeuAsnLeuA 945
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
4158 AGGTTTC.....CAGAAGCTTGAGAACCTTAAGCTTT 4189
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
945 laGIyAsn...ArgValSerSerAspGIyTrpLeuAlaPheMetGIyVal 960
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
4190 CAATTCATATCCAAAGATTACAGAGAGATACAGAATTTCTTCAAGCA 4239
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
961 PheGIuAsnLeuLysGIuLeuValPhePheAspPheSer.....Th 974
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
4240 CTGGACACACATGGCAAACTTGCAGAGTTGACATCTCCAGGCATTTTAC 4289
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
974 rIySGIuPheLeuProAspProAlaLeuValArgGIyLysSerGIuVal 991
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
4290 AGAGGTATCAAAAGCTCAGGCCACACAGTCAAGTCTTTGATCAATGTG 4339
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
991 eueSerLysLeuThrPheLeuGIuGIuAlaArgLeuValGIyTrpGIuPhe 1007
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
4340 TGTAGACATACCAAGGCTCATTAAGTCAACATGTTAAGTTGCGCTTG 4389
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1008 AspAspAspAspLeuSerValIle 1015
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
4390 GATGCAGATGATATGATGCTT 4413
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
seq_name: gb_pat:A64529

```

```

AUTHORS Korneluk,R.G., Mackenzie,A.E., Roy,N., Robertson,G. and Tamai,K.
TITLE USE OF NEURONAL APOPTOSIS INHIBITOR PROTEIN (NAIP)
JOURNAL Patent: WO 9726331-A 21 24-JUL-1997;
UNIV OTTAWA (CA)
COMMENT Other Publication AU 1614997 19970811.
FEATURES Location/Qualifiers
source 1. 6124
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 1803 a 1298 c 1285 g 1738 t
ORIGIN
alignment_scores:
Quality: 653.00 Length: 1108
Ratio: 1.074 Gaps: 42
Percent Similarity: 54.874 Percent Identity: 23.736
alignment_block:
us-09-697-089-2 x A64529 ..
Align seg 1/1 to: A64529 from: 1 to: 6124
69 PheLeuLysSerLeuLysGIuTrpAsnTyrProLeuPheGlnAspLeuS 85
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1321 TTTCTCCAAATATGAAGTCC...TCTGCGAGAGTACTCCACACCTTCA 1367
85 ngIyGln.....SerLeuPheHisGIuThrSerGIyGIyAspL 98
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1368 GAGCCGGGTGAACCTTTGTGAATTACTGCAACCAAGCAAGCAAGCAATC 1417
98 euaSpAsp..... 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1418 TTGAAGATTCAATGACAGTTGCTCTATAGTCCAGCAATGCGCAGAGGT 1467
101 .....LeuAlaGlnAspLeuLysAs 107
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1468 GAAGCCAGTGTTCAGAGCGCAAGAAATCTGAATGAGCAGCTGAGAGC 1517
107 pleuTyrHisThrProSerPheLeuAsnPhe..... 117
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1518 AGCTTATACAGCGCGCATTTCCGCCACATGCTTTCCTGATATCTCTT 1567
118 .....TyrProLeuGIyGIuAspIleAspIlePhe 128
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1568 CCGATCTGCGCCAGCAGCACCTTCTGCGGTGATGCTGTATTT..... 1611
129 AsnLeuLysSerThrPheThrGIuProValLeuTrpArgLysAspGlnH 145
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1612 ..GCTTCAAAACATACAGCAACCTGTG..... 1638
145 shIshIshArgValGIuGIuLeuThrLeuAsnGIyLeuLeuGlnAlaLeu 162
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1639 .....CAAGAACCTTGTGCTGCTGCCCTGAGCTTTGGCAACCTTGA 1678
162 InseProCysIleIleGIuGIuGIuSerGIyLysGIyLysSerThrLeu 178
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1679 ACTGTGTATGTGTGTGAGGAGGTGAAGCTGGAAGTGAAGAGAGGTCCTC 1728
179 LeuGlnArgIleAlaMetLeuTrpGIySerGIyLysCysLysAlaLeuTh 195
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1729 CTGAAGAAATACCTTTCTGTGGCATGTGATGCTGCTCCCTGTATAA 1778
195 rIySGIuPheLysPheValPhePheLeuArgLeuSer.....ArgAlaGln 210
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1779 CAGGTTCCAGCTGTGTTTACCTCTCCCTAGTTACACAGACACAGAGAG 1828
210 LysGIyLeuPheGIuThrLeuCysAspGIuLeuLeuAspIleProGIyLyr 226
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1829 AGGGGCTGGCCAGTATCATCTGTGACCAAGCTCTTAAGAAAGAGAGATCT 1878
227 IleArgLysGlnThrPheMetAlaMetLeuLeuLysLeuArgGlnArgVa 243
|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

1879 GTTACGAAAGTCTGCATGAGAACATTATCCAGACAGTTAAAGATCAGT 1928
243 LLeuPheLeuLeuAspGlyTyrAsnGluPhe.....LysProGlnAsnC 258
||||| ||||| ||||| |||||
1929 CTTATCTCTTTAGATGACTACAAAGAAATATGTTTCATCCCTCA... 1974
258 ysrProGluIleGlnAlaLeuIleLysGlnAsnHisArgPheLysAsnMet 274
||||| ||||| ||||| |||||
1975GTCATAGAAAACGTGATTCAAAAAAAACCATTAATCCCGAGCTGC 2019
275 ValIleValThrThrThrThrGluCysLeuArgHisIleArgGlnPheG1 291
||||| ||||| ||||| |||||
2020 CATTGATGTCGTCTCCGATACAAACAGGGCCAGGGACATCCGCCATATCT 2069
291 yAlaLeuThrAlaGluAlaGlyAspMetThrGluAspSerAlaGlnAla 308
||||| ||||| ||||| |||||
2070 AAGAGACATTTCTAGAGATCAAGCAATTTCCCTTTATATATCTGTCTGA 2119
308 euIIleArgGluValLeuIleLysGluLeuAla.....GluGlyLeuLeu 322
||||| ||||| ||||| |||||
2120 TATTACGGAAGCTCTTTACATATATAGACTGCTGCGCAAGTTTATG 2169
323 LeuGlnIleGlnLysSerArgCysLeuArgAsnMetLysThrProle 339
||||| ||||| ||||| |||||
2170 GTTTACTTTGGAAAGAACCAAGTTTGCGAGAGATACAGAAACTCTCT 2219
339 uPheValAlaIleThrCysAlaIleGlnMetGlyLysSerGluPheHis 356
||||| ||||| ||||| |||||
2220 CTTTGGGGGGGAGATCTGCGCATGTGTTGTCAGATCTTTGACCAT 2269
356 eThIstHsgIlnThrThrLeuPheHisThrPheIlyrAspLeuLeuIleGln 372
||||| ||||| ||||| |||||
2270 CTTTGAATGATGTGGCTCTTTTCAAGTCCATATGCAAGCGCTTTCTTA 2319
373 LysAsnLysHisLysHisLysGlyValAlaIleAspArgPheIleArg.. 388
||||| ||||| ||||| |||||
2320 AGGAACAAA.....GCGACAGTGAATTTCTCAAGC 2351
389 ..SerLeuAspHisCysGlyAspLeuAlaIleGluGlyValPheSerHisL 405
||||| ||||| ||||| |||||
2352 AACTGCTGCTCTGCTGGAGAGCTGCTTGAAGGGTTTTTTTCATGTT 2401
405 ysrPheAspPheGluLeuGlnAspVal.....SerSerValAsnGluAsp 419
||||| ||||| ||||| |||||
2402 GCTTTGAGTTTATATATGATGATGATCCGACGAAGAGGGCTGATGAAGAT 2451
420 ValLeuLeuThrThrGlyLeuLeuCysLysTyrThrAlaGlnArgPheIly 436
||||| ||||| ||||| |||||
2452 GAAGATCTACCATGCTGCTGATGAGCAAAATTTACAGCCACAGACTATAG 2501
436 sProlLysTyrLysPhePheHisLysSerPheGlnGluThrAlaGly 453
||||| ||||| ||||| |||||
2502 ACCATTTCTACCGGTTTTTAAAGTCGCTTCCAAAGAAATTTCTTGCGGGA 2551
453 rArgLeuSerSerLeuLeuThrSerHisGluProlGluGluValThrLys 469
||||| ||||| ||||| |||||
2552 TCAGAGCTGATTCGAATCTCGGATTCAGATACGAGCAAGATCAAGATTGG 2601
470 GluAsnGlyTyrLeuGlnLysMetValSerLLeuSerPheThrSerTh 486
||||| ||||| ||||| |||||
2602 GGACTGTATCATTTTAAACAAATCAACACACCCTGATGATGATGACGCG 2651
486 rTyrSerSerLeuLeuArgTyrThrCysGlySerSerValGluAlaThr 503
||||| ||||| ||||| |||||
2652 CTACACAAATTTTGGAACTATGTC.....TTCAGGCTCCCTTCAACAA 2695
503 rGAla.....ValMetLysHisLeuAlaAlaValTyrGlnHisGly 516
||||| ||||| ||||| |||||
2696 AAGCAGGGGCGCAAAATATGCTGTATTTGTCATTTAAGTGAGATAAC... 2742
517 CysLeuLeuGlyLeuSerIleAlaLysArgProlLeuThrArgGlnGluSe 533
||||| ||||| ||||| |||||
2743AAGAGTCT 2750

533 rleuGlnSerValIlyIysAsn..... 539
 2751 ATTTGGAGGAATATATCTGAAAAATGATGACTACTAAAGCACAGCCAGCAA 2800
 540 ... ThrThrcIuInGlnIuLeuIysAlaIle..... AsnIle 551
 2801 TTTCACGTGAGATGCAGTTACTTACGGGATGTGTGGCAATTGTGCCACA 2850
 552 AsnSerPheValGluCysGlyTleHisLeu..... 561
 2851 GCTTACTTTTCAMTGGTTTCAGAACATTTACTGTCTTCCCTCGAANAAC 2900
 562 ... TyrGlnGlnSerThrSerLysSerAlaLeuSerGlnIuPheGln 577
 2901 TGGTATATCAAGACACACT... GTTGGCGGTGTCTTCCACTTTGTGTTGC 2947
 577 IAPhePheGlnGlnIlyLysSerLeuTyrIleAsnSerGlyAsnIleProAsp 593
 2948 AATTCCCTTCACAGGAGACACTGACTTGTGGTGGCTTAACCTTA... CAG 2994
 594 TyrLeuPheAsp..... 597
 2995 TACTCTTTCGACCCACCAAGACTTGTCTATTTGTGAGAGACATCCACTT 3044
 598 PhePheGlnIuHisLeuP 603
 3045 CCCAATATACAGAGAATATAGACATCATCCACGACGACACATTTTTCAGTTCTGG 3094
 603 roAsnCys..... AlaSerAlaLeuAspPheIleLysLeuAspPheTyr 617
 3095 AAACATGCTTTTGACAAATCACAGGCGCCAACTATATGATCAGGACTATAGCT 3144
 618 GlyGly..... AlaMetAlaSerTrpGlu..... 625
 3145 TCTGCCCTTGAACCTATATGATGATGAGGAGCGCAATTAGCTGAAAAAGA 3194
 626 LysAlaAlaGlnAspT 631
 3195 GGATATATGTAAGAGACTATATGATATGACAGCGCAGGCGATCACAGCAAGCC 3244
 631 hGlyGlyLysIleHisMetGlnIuAlaProGluTyrTrp... IleProSer 646
 3245 TTAGTACGTGGCATTTGGAAGACTTCTCCAAAGCAGTACCAAGATTCCC... 3291
 647 ArgAlaValaSerLeuPheAsnTrpLysGlnIuPheArgTyrLeuGln 663
 3292 TGTCTAGA 3299
 663 uValAlaThrLeuArgAspPheSerLysLeuAsnLysGlnAspIleThrTyr 680
 3300 AGTCGATGGAATGATATGATGATTGTGAGCCAGGATATGCTTGAGATTTC 3349
 680 GlnGlyLysIlePheSerSerAlaThrSerLeuAlaGlnGlnIleLysArg 696
 3350 TATAGCAGATTTCTCCACTTCACAGCGCAGCAATCGAATCCATTTAAACCAAC 3399
 697 CysAlaGlyValaAlaGlySerLeuSerLeuValaLeuSerThrcysLys.. 712
 3400 AGCAGAGGCTTTATGAAAGCATCGGCCCACTCTTGAGCTGTCTAAGGC 3449
 713 AsnIleTyrSerLeuMetValGlnAlaSerProLeuThrIleGlnAspG 729
 3450 CTCTGTACCCAAAGTCTCCATTAAGCAAGTTGGAACTCAAGGCAAGCCGAAAC 3499
 729 IuArgHisIleThrSerValaThrAsnLeuLysTrpThrSerIleHis... 744
 3500 AGGAACGTGTTCTACACCTGCGCTCCGCGAATCTCTTGAAGTCOTAGGG 3549
 745 AspLeuGlnAsnGlnArgLeuProGlyGlyLeuThrAspSerIle 759
 3550 ACAATTCACATTCACAGCAAGCAATCTTCTCATAT... CTGATGATAGTT 3593


```

: 389 .SerLeuAspHisCysGlyAspLeuAlaLeuGluGlyValPheSerHisL 405
:      :      :      :      :      :      :      :      :      :
2352 AACTGTCTCTCCCTGTGAGCTGGCCCTGAAGAAGGCTTTTTCATGTT 2401
405 ysrPheAspPheGluLeuGlnAspVal.....SerSerValasnGluAsp 419
:      :      :      :      :      :      :      :      :      :
2402 GCCTTGTAGTTTAATGATGATGATCTGCACAGACAGGCGTTGATGAAGAT 2451
420 ValLeuLeuThrThrGlyLeuLeuCysLysThrThrLagLlnArgPheL 436
:      :      :      :      :      :      :      :      :      :
2452 GAAGATCTAACCATGCTGTGATGAGCAAAATTACACCCAGACACTAAG 2501
436 sProLysThrLysPhePheHisLysSerPheGlnGluThrThrAlaGlyA 453
:      :      :      :      :      :      :      :      :      :
2502 ACCATTCTACCGGTTTTTAAGTCTCGCTCCAAAGATTCTTGGCGGGA 2551
453 rGArgLeuSerSerLeuLeuThrSerHisGluProGluValThrLys 469
:      :      :      :      :      :      :      :      :      :
2552 TGAGGCTGATGAACCTCGGATTGAGATGAGCAGACATCAAGATTGG 2601
470 GlyAsnGlyThrLeuGlnLysMetValSerLiesAspLierHisSerTh 486
:      :      :      :      :      :      :      :      :      :
2602 GGACTGTATCATTTGAACAACAACTACCCCATGATGACTGTAAAGCGC 2651
486 rTyrSerSerLeuLeuArgThrThrCysGlySerSerValGluAlaThrA 503
:      :      :      :      :      :      :      :      :      :
2652 CTACAAACAATTTTGAAGTATGTC.....TCCAGCTCCCTTCAACA 2695
503 rGAla.....ValMetLysHisLeuAlaValAlaGlyGlnHisGly 516
:      :      :      :      :      :      :      :      :      :
2696 AACACAGGCCCAAAATGTGTCTCATTTGTCCATTAGTGAATAC... 2742
517 CysLeuLeuGlyLeuSerLiesAlaLysArgProLeuThrArgLlnGlu 533
:      :      :      :      :      :      :      :      :      :
2743 .....AAAGAGTC 2750
533 rLeuGlnSerValLysAsn..... 539
:      :      :      :      :      :      :      :      :      :
2751 ATTGGAATATATCTGAATAATGATGACTACTTAAGCACACCACAGAAA 2800
540 .ThrThrGlnGlnGluLeuLysAlaLe.....AsnIle 551
:      :      :      :      :      :      :      :      :      :
2801 TTTCACGTGACGATGACATTACTAGGGAATTGTGCACAAATTGTCCACA 2850
552 AsnSerPheValGluCysGlyLlnHisLeu..... 561
:      :      :      :      :      :      :      :      :      :
2851 GCCTTACTTTCAATGCTTTCAGAACATTTACTGTTCTTGCCTGAAAC 2900
562 ....TYGlnGlnLysThrSerLysSerAlaLeuSerGlnGluPheGluA 577
:      :      :      :      :      :      :      :      :      :
2901 TGGTTATCAAGACACT...GTGCTGGGTGTCTCCATTTGTTTGC 2947
577 lAherPheGlnGlyLysSerLeuThrLlLeasnSerGlnLlLeuAsp 593
:      :      :      :      :      :      :      :      :      :
2948 AATTCTCTCAAGGAGACACTGACTTTGGTGCCTTAACCTTA...CAG 2994
594 TyrLeuPheAsp..... 597
:      :      :      :      :      :      :      :      :      :
2995 TACTTTTTCAGCACCCAGAACCTTGTCACTTGTGAGAGCATCCACTT 3044
598 .....PhePheGlnLlnHisLeuP 603
:      :      :      :      :      :      :      :      :      :
3045 CCCAATACGAGAAATAGACATCACCCAGACACATTTTTCAGTTCTGG 3094
603 roAsnCys.....AlaSerAlaLeuAspPheLlLeuLysLeuAspPheTyr 617
:      :      :      :      :      :      :      :      :      :
3095 AAACATGTTTTCACAATACACAGTGCACAACTATGATCAGGACTATGCT 3144
618 GlyGly.....AlaMetAlaSerTrpGlu..... 625
:      :      :      :      :      :      :      :      :      :
3145 TCTGCCCTTAACTATGAAATGAGGAGACGAAATTTAGCTAAAAAGA 3194
626 .....LysAlaAlaGluAspTr 631

```

```

3195 GCATTAATGTAAAGACCTATATGATATGCAGCCAGGACATCACAGACC 3244
:      :      :      :      :      :      :      :      :      :
631 hrcLysGlyLlnHisMetGluGluAlaProGluThrTyr...lIleProSer 646
:      :      :      :      :      :      :      :      :      :
3245 TTAGTACTGGCTATTTGAAAACCTTCTCAAGCAGTACAAAGATCC... 3291
647 ArgAlaValSerLeuPheAsnTrpLysGlnGluPheArgThrLeuGlu 663
:      :      :      :      :      :      :      :      :      :
3292 .....TGTTTAA 3299
663 uValThrLeuArgAspPheSerLysLeuAsnLysGlnAspLierThrTrL 680
:      :      :      :      :      :      :      :      :      :
3300 AGTCGATGTAATGATATGATGTGTAGGCCAGATATGCTTGACATTC 3349
680 euGlyLysLlIePheSerSerAlaThrSerLeuArgLeuGlnlIleLysArg 696
:      :      :      :      :      :      :      :      :      :
3350 TAATGACACTTTTCTCAGCTTCACACCGCATCGAACCTCATTTAAACAC 3399
697 CysAlaGlyValAlaGlySerLeuSerLeuValLeuSerThrCysLys.. 712
:      :      :      :      :      :      :      :      :      :
3400 AGCAGAGCCTTTATAGAAAGCATCCGCCACTTGTGACGTCTTAAGCC 3449
713 .AsnIleTyrSerLeuMetValGlnAlaSerProLeuThrLlGluAspG 729
:      :      :      :      :      :      :      :      :      :
3450 CTCTGTACACCAAGTCTCCATTAAGCAAGTTGGAACCTCAGCGACCGGAC 3499
729 lAArgHisLlIeThrSerValThrAsnLeuLysThrLeuSerLlnHis... 744
:      :      :      :      :      :      :      :      :      :
3500 AGGAACGTCTTCACACCTCCCTCGGAACTCTGTGAAGCTCAGCG 3549
745 .....AspLeuGlnAsnGlnArgLeuProGluGlyLeuThrAspSerLe 759
:      :      :      :      :      :      :      :      :      :
3550 ACAATCCAGTCACAGACCAATCTTCCATAT.....CTGATTAAGTT 3593
759 uGlyAsnLeuLysAsnLeuThrLysLeuLlIleMetAspAsnLlIleLysMet 775
:      :      :      :      :      :      :      :      :      :
3594 CCTGTCCCTGAAGAAGACCTGTGTGATCTGGAGGCACTATATAAGTTT 3643
775 ..... 775
3644 TTTCACGTATCTCTGAAGAATTTCCAAACCTCCACCATGAGAGAAATTA 3693
776 .....AsnGluGluAspAlaLlIleLysLeuAlaGluGlyLe 787
:      :      :      :      :      :      :      :      :      :
3694 TTGATCCAAATTTACCTGATGATGATCTCCAACTAGTAAATTAAT 3743
787 uLysAsnLeuLysLysMetCysLeuPheHisLeu.....ThrHisLeuS 802
:      :      :      :      :      :      :      :      :      :
3744 TCAAATTTCTCCAAACCTTCATGTTTCCATCGAAGTAACTTCTTT 3793
802 eTrAspLlIeGlyGlnGlyMetAspTrpLlIleValLysSerLeuSerGlu 818
:      :      :      :      :      :      :      :      :      :
3794 CGGATTTTGGTCTCATGACTATGCTGTTC..... 3828
819 ProCys...AspLeuGlnGluLlIeGlnLeuValSerCysCysLeuSerAl 834
:      :      :      :      :      :      :      :      :      :
3829 ...TGTAAACACTCACAGAAATTAAGTTTTCGATTCATTTT... 3870
834 AsnAlaValLysLlIleLeuAlaGlnAsnLlnHisAsnLeuValLysL 851
:      :      :      :      :      :      :      :      :      :
3871 .CAAGCGTCCCATTTGTGGC...AGTTGCAAAATTTATTTCTGTGA 3916
851 eTrLlIeLeuAspLeu...SerGluAsnTrpLeuGlnLulLysAspLysGln 866
:      :      :      :      :      :      :      :      :      :
3917 AGATATTAATCTTTGAAGCCAGCAATTTCTGTATGAGGAAACATCAGAA 3966
867 AlaLeuHisGluLeuLlIeAspArgMetAsnValLeuGlnGlnLeuThrAl 883
:      :      :      :      :      :      :      :      :      :
3967 AATTGTGCTTACATTTTAGCTTCTGTACTAACCTGGAGGAA..... 4008
883 AlaMetLeuProTrpGlyCysAspValGlnGlySerLeuSerSerLeuL 900
:      :      :      :      :      :      :      :      :      :

```

```

4009 .TTGATCTTCTACTGGGATGGAATTATTCAGCTGGCCAAATGATCA 4057
900 eulysH1sleuGlulValProGlnleuValIlysleuGlyleuLysasn 916
      ::::: ::::: ||| ::::: :::::
4058 TCCAGCATGTCAGCAGCTTCATTCTCTCCAGTCTCTCATTTTTCAG 4107
917 TTPATgLeuThrAspThr.....GluIleArgIleLeuG1 928
      ||::: ||::: ||::: ||:::
4108 ACTTTGATGATGACAGCGGTGGTGAATTCACAAATGACAAATCAGTGG 4157
928 yAlaPhePheGlyLysAsnProLeuLysAsnPheGlnIleuAsnLeuA 945
      ::::: ||::: ::::: ||::: :::::
4158 AGGTTTC.....CAGAACTTTCAGAACTTAAAGCTTT 4189
945 lAglyAsp...ArgValSerSerAspGlyTrpLeuAlaPheMetGlyVal 960
      ::||| ::::: ::::: ||::: ||:::
4190 CAATCAATACACAAGATTACAGAGAGATACAGAAATTTCTTTCAACCA 4239
961 PheGluAsnLeuLysGlnLeuValPhePheAspPheSer.....Th 974
      ::::: ||::: ::::: ||::: :::::
4240 CTGGACACATGCCAAACTGCAGAGATTGGACATCTCCAGCATTTTCAC 4289
974 rLysIleuPheLeuProAspProAlaLeuValArgLysLeuSerGlnVal 991
      ||::: ::::: ||::: ||::: ||::: ||:::
4290 AGAGTGTATCAAAAGCTCAGGCCACACACAGTCAGCTTTGAGTCATGTG 4339
991 euSerLysLeuThrPheLeuGlnGlnAlaArgLeuValGlyTrpGlnPhe 1007
      ::::: ||::: ||| ::::: ::::: ||:::
4340 TGTTCAGCATACCAAGCTCATTTAGACAGTCAACATGTTAAGTTGGCTTG 4389
1008 AspAspAspLeuSerValIle 1015
      ||| ||||| ::::: :::::
4390 GATCGAGATGATATGCAATGCTT 4413
seq_name: gb_pat:A64510

seq_documentation_block:
LOCUS A64510 6133 bp DNA PAT 29-MAR-1999
DEFINITION Sequence 2 from Patent WO9726331.
ACCESSION A64510
VERSION A64510.1 GI:3717909
KEYWORDS
SOURCE
  unidentified.
  unidentified.
  unclassified.
  1 (bases 1 to 6133)
REFERENCE
  KORNEIUK,R.G., MACKENZIE,A.E., ROY,N., ROBERTSON,G. and TAMAI,K.
  TITLE USE OF NEURONAL APOPTOSIS INHIBITOR PROTEIN (NAIP)
  JOURNAL Patent: WO 9726331-A 2 24-JUL-1997;
  UNIT OTTAWA (CA)
  OTHER PUBLICATION AU 1614997 19970811.
  LOCATION/Qualifiers
  source 1. 6133
  /organism="unidentified"
  /db_xref="taxon:32644"
BASE COUNT 1806 a 1299 c 1288 g 1740 t
ORIGIN

```

```

alignment_scores:
  Quality: 653.00 Length: 1108
  Ratio: 1.074 Gaps: 42
Percent Similarity: 54.874 Percent Identity: 23.736

```

```

alignment_block:
US-09-697-089-2 x A64510 ..

```

```

Align seg 1/1 to: A64510 from: 1 to: 6133

```

```

69 PheLeuLysSerLeuLysGluTrpAsnThrProLeuPheGlnAspLeuAs 85
|||||:::|||||::: |||::: |||||:::
1330 TTTCTCCAAATATGAAAGTCC...TCTCGGAGAGTACTCCAGACCTTCA 1376

```

```

85 nGlyIln.....SerLeuPheHisGlnThrSerGluLysPRL 98
      ::::: ||::: ||::: ||::: ||::: ||:::
1377 GAGCGGTGTGTAACCTTGTGTAATTACTGAAACCAACCAAGTAAAGCAATC 1426
98 euAspAsp..... 100
      ||::: ||::: ||::: ||::: ||:::
1427 TTGAAGATTCAATAGCAGTTGCTCTATATGTCCAGAAATGCGCAGAGGT 1476
101 .....LeuAlaGlnAspLeuLysAs 107
      ||| ::::: ||::: ||::: ||:::
1477 GAAGCCCGAGTGTTCACAGAGCAAGACATCTGAATGAGCGTGAAGAC 1536
107 PLeuThrHisThrProSerPheLeuAsnPhe..... 117
      ||| ::::: ||::: ||::: ||:::
1527 AGCTTATTCACAGCGCAGTTTCCGCCACATGCTTTGCTGATATCTCTT 1576
118 .....TyrProLeuGlyGluAspIleAspIlePhe 128
      ::::: ||::: ||::: ||::: ||:::
1577 CCGATCTGGCCACGACCGACCTGCTGCTGTGATCTGTATTT..... 1620
129 AsnLeuLysSerThrPheThrGluProValLeuThrArgLysAspGlnI 145
      ::::: ||::: ||::: ||::: ||:::
1621 ..GCTTAAACACATCAGCAAACTGTG..... 1647
145 sHisHisArgValGlnGlnLeuThrLeuAsnGlyLeuGlnAlaLeuG 162
      ||| ||::: ||::: ||::: ||:::
1648 .....CAAGAACCTGTGTGCTGCTGAGGTCTTTGGCAACTTGA 1687
162 InSerProCysIleIleGlnGlyLysSerGlyLysGlyLysSerThrLeu 178
      ::::: ||::: ||::: ||::: ||::: ||:::
1688 ACTCTGTCATGTGTGTGAGGAGTGAGCTGAGAGTGAAGAGCAAGCGTCTC 1737
179 LeuGlnArgIleAlaMetLeuThrIleGlySerGlyLysCysLysAlaLeu 195
      ||::: ||::: ||::: ||::: ||::: ||:::
1738 CTGAAGAATAATACCTTTCTCTGCGGCAATCTGATGCTGCTCCCTGTTAA 1787
195 rLysPheLysPheValPhePheLeuArgLeuSer.....ArgAlaGlnG 210
      ::::: ||::: ||::: ||::: ||::: ||:::
1788 CAGGTTCCAGCTGTTTCTACTCTCCCTTATGTTCCACAGCACAGCAGAG 1837
210 LysGlyLeuPheGluThrLeuLysAspGlnLeuLeuAspIleProGlyThr 226
      ||||| ::::: ||::: ||::: ||::: ||::: ||:::
1838 AGGGCTGTGGCCAGTATCATCTGTGACAGCTCTGAGAAAGAAAGAGATCT 1887
227 IleArgLysGlnThrPheMetAlaMetLeuLeuLysLeuArgGlnArgVa 243
      ::::: ||::: ||::: ||::: ||::: ||:::
1888 GTTACTGAAATGTGCATCAGACATATATCCAGCAGTTAAAGATTCAGGT 1937
243 lLeuPheLeuLeuAspGlyTyrAsnGluPhe.....LysProGlnAsn 258
      ||||| ||::: ||::: ||::: ||::: ||:::
1938 CTATATTCCTTTTATAGTACTCAAAAGAAATATGTTCAATCCCTCA... 1983
258 ySPROGlnIleGlnAlaLeuIleLysGlnAsnHisArgPheLysAsnMet 274
      ||| ||||| ::::: ||::: ||::: ||:::
1984 ....GTCATGTGAAAACTGATTCAAAAAAACCACTTATCCCGGACCTGCG 2028
275 ValIleValThrThrThrThrGluCysLeuArgHisIleArgGlnPheG1 291
      ::::: ||::: ||::: ||::: ||::: ||:::
2029 CTATTTGATTGCTGTCCGTACAAACAGGCGCCAGGACATCCGCCATACCT 2078
291 yAlaLeuThrAlaGluValGlyAspMetThrGluAspSerAlaGlnAla 308
      ||::: ||::: ||::: ||::: ||::: ||:::
2079 AGAGACCATTTCTAGAGATCAAAAGCATTTCCCTTTATATATCTGTGTA 2128
308 euIleArgGluValLeuIleLysGluLeuAla.....GluGlyLeuLeu 322
      ::::: ||::: ||::: ||::: ||::: ||:::
2129 TATTACGGAAGCTCTTTTCACATATATAGACTGCTGCGAAAGCTTATG 2178
323 LeuGlnIleGlnLysSerArgCysLeuAsnLeuMetLysThrProLe 339
      ::::: ||::: ||::: ||::: ||::: ||:::
2179 GTTACTTGTGAAAGAACCAAGATTGTGAGAAAGATACGAAAAACTCTCT 2228
339 uPheValValIleThrCysAlaIleGlnMetGlyLysSerGluPheHis 356

```

```

2229 CTTTGTGGCGGCGATCTGCTCATTTGGTTTCAGTATCCTTTGACCATT 2278
      |||||::: |||||:::
356 erh1sthrnglnthrleupheh1sthrphetyraspleu1le1gln 372
      ||:::|||||:::|||||:::
2279 CCTTGTATGATGATGGCTTTTCAAGCTATATGAAAGCGCTTCCCTTA 2328
      ||:::|||||:::|||||:::
373 LysAsnLysHisLysHisLysGlyValAlaAlaSerAspHe1learg 388
      ||:::|||||:::|||||:::
2329 AGGACAAA.....GGCAGACGTGAATTCTCAAGC 2360
      ||:::|||||:::|||||:::
389 ..SerLeuAspHisCysGlyAspLeuAlaLeuGlnGluValPheSerHisL 405
      ||:::|||||:::|||||:::
2361 AACTGATGCTCCTGCTGGTGGAGCTGGCTGAAGGCTTTTTCATGTT 2410
      ||:::|||||:::|||||:::
405 yspheAspHe1glnLeuGlnAspVal.....SerSerValAsnGluAsp 419
      ||:::|||||:::|||||:::
2411 GCCTTGAATTAAATGATGATGATCGCAGAACAGCGGCTTGATGAAGAT 2460
      ||:::|||||:::|||||:::
420 ValLeuLeuThrThrGlyLeuLeuCysLysSerThrAlaGlnArgPheL 436
      |||||::: |||||:::
2461 GAAGATCTAACCATGCTGCTGATGAGCAAAATTTACAGCCAGAGACTAAG 2510
      |||||::: |||||:::
436 sProLysTyrLysPhePheHisLysSerPheGlnGluTyrThrAlaGly 453
      |||||::: |||||:::
2511 ACCATCTACCGGCTTTTAAGTCTGCTGCCCAAGAAATTTCTGGCGGGA 2560
      |||||::: |||||:::
453 rgrArgLeuSerSerLeuLeuThrSerHisGluProGlnGluValThrLys 469
      |||||::: |||||:::
2561 TGAAGCGGATGAACTCGTGATTCAGATAGGAGCAAGACATCAAGATTGG 2610
      |||||::: |||||:::
470 GlyAsnGlyTyrLeuGlnLysMetValSerIleSerAspIleThrSerTh 486
      |||||::: |||||:::
2611 GGACTGATGATATTGAAACAATCAACTCCACCATGATGACTGTAAGCCG 2660
      |||||::: |||||:::
486 rTyrSerSerLeuLeuArgTyrThrCysGlySerSerValGluAlaThrA 503
      |||||::: |||||:::
2661 CTACACAAATTTTGTGAATATGTC.....TCCAGCCTCCCTTCAACAA 2704
      |||||::: |||||:::
503 rglAa.....ValMetLysHisLeuAlaAlaValTyrGlnHisGly 516
      |||||::: |||||:::
2705 AACGAGGCCCAAAATTTGTGTCATTTGCTCCATTGATGAGTAAC... 2751
      |||||::: |||||:::
517 CysLeuLeuGlyLeuSerIleAlaLysArgProLeuTyrArgGlnLys 533
      |||||::: |||||:::
2752 .....AAAGAGTC 2759
      |||||::: |||||:::
533 rLeuGlnSerValLysAsn..... 539
      |||||::: |||||:::
2760 ATTGGAGAAATATATCTGAATAATGATGACTACTTAAAGCACACAGCAAA 2809
      |||||::: |||||:::
540 ..ThrThrGlnGlnGluIleLeuLysAlaIle.....AsnIle 551
      |||||::: |||||:::
2810 TTTCACTGCGAGATGACAGTACTTAGGGGATTTGGCAAAATTTGTCCAAA 2859
      |||||::: |||||:::
552 AsnSerPheValGluCysGlyIleHisLeu..... 561
      |||||::: |||||:::
2860 GCTTACTTTTCATGCTTTCAGAACATTTACTGCTTCCCTCCCTGAAGC 2909
      |||||::: |||||:::
562 ....TyrGlnGluSerThrSerLysSerAlaLeuSerGlnGluPheGlu 577
      |||||::: |||||:::
2910 TCGCTTATCAAGCAACACT...CTTGCTGCGTGTCTCCATTGTGTTTGC 2956
      |||||::: |||||:::
577 lAhePheGlnGlyLysSerLeuTyrIleAsnSerGlnLysAsnIleProAsp 593
      |||||::: |||||:::
2957 AATTCCCTTCAAGGAGAACACTGATTTGGTGCGCTTAACCTTA...CAG 3003
      |||||::: |||||:::
594 TyrLeuPheAsp..... 597
      |||||::: |||||:::
3004 TACTTTTTCACACACCCAGAAAGCTTGATCTTTGTGAGAGCATCCACTT 3053
      |||||::: |||||:::
598 .....PhePheGlnHisLeuP 603
      |||||::: |||||:::
3054 CCCAATACGAGAAATTAACATCATCCACGAGCACAATTTTCAGTCTGG 3103
      |||||::: |||||:::
603 roAsnCys.....AlaSerAlaLeuAspPheIleLysLeuAspPheTyr 617
      |||||::: |||||:::
3104 AAACATGCTTTTGAACAATCACAGGTGCCAACATTAATGATGAGACTATGCT 3153
      |||||::: |||||:::
618 GlyGly.....AlaMetAlaSerTrpGlu..... 625
      |||||::: |||||:::
3154 TCTGCCCTTGAACCTTATGAATGAATGAGACCGAAATTTAGCTAAAAAGA 3203
      |||||::: |||||:::
626 .....LysAlaAlaGluAspT 631
      |||||::: |||||:::
3204 GGATTAATGTAAGACGATATATGATATGACAGCGGCGCATCCAGAGCC 3253
      |||||::: |||||:::
631 hrGlyGlyIleHisMetGlnGluAlaProGlnThrTyr...IleProSer 646
      |||||::: |||||:::
3254 TTGATCTGCGCTATTTGAAACTTTTCTCCAAAGCAGTACAGATTTCC... 3300
      |||||::: |||||:::
647 ArgAlaValSerLeuPhePheAsnTrpLysGlnGluPheArgThrLeuG 663
      |||||::: |||||:::
3301 .....TGCTTAGA 3308
      |||||::: |||||:::
663 uValThrLeuArgAspPheSerLysLeuAsnLysGlnAspIleThrTyrL 680
      |||||::: |||||:::
3309 AGTCGATGTAATGATATGATGATGTTAGGCCAGATATGCTTGAGATTC 3358
      |||||::: |||||:::
680 euGlnLysIlePheSerSerAlaThrSerLeuArgLeuGlnIleLysArg 696
      |||||::: |||||:::
3359 TAATGACAGTTTCTCAGCTTCACAGCGCATCGAATCTCATTTAAACAC 3408
      |||||::: |||||:::
697 CysAlaGlyValAlaGlySerLeuSerLeuValLeuSerThrCysLys... 712
      |||||::: |||||:::
3409 AGCAGAGCGCTTTATAGAAGCATCCGCCACGCTTGAGCTGTCTAAGGC 3458
      |||||::: |||||:::
713 ..AsnIleTyrSerLeuMetValGluAlaSerProLeuThrIleGluAsp 729
      |||||::: |||||:::
3459 CTCTGTACACCAAGTCTCCATTAAGCAAGTTGGAATCAGCGCAGCCGAA 3508
      |||||::: |||||:::
729 LuArgHisIleThrSerValThrAsnLeuLysThrLeuSerIleHis... 744
      |||||::: |||||:::
3509 AGGAACTGCTTCACCCCTGCTCCCTCGGAATCTCTTGAAGCTCAGGG 3558
      |||||::: |||||:::
745 .....AspLeuGlnAsnGlnArgLeuProGlnGlyLeuThrAspSerLe 759
      |||||::: |||||:::
3559 ACAATCCAGTCACAGACCAATCTTCCAT...CTGATTAAGTT 3602
      |||||::: |||||:::
759 uGlyAsnLeuLysAsnLeuThrLysLeuIleMetAspAsnIleLysMet. 775
      |||||::: |||||:::
3603 CCTGTCCCTGAAAGACTGTCTGTGATCTGGAGGCAATATTAATGTTT 3652
      |||||::: |||||:::
775 ..... 775
      |||||::: |||||:::
3653 TTTCACTCATTTCTGAAAGATTTCCAAACTTCCACCATATGGAATAATTA 3702
      |||||::: |||||:::
776 .....AsnGlnGluAspAlaIleLysLeuAlaGlnGlyLe 787
      |||||::: |||||:::
3703 TTGATCCAAATTTCACTGATGATGATCCCTCCAACTAGTAATAATTAAT 3752
      |||||::: |||||:::
787 uLysAsnLeuLysLysMetCysLeuPheHisLeu.....ThrHisLeus 802
      |||||::: |||||:::
3753 TCAAAATTTCTCCAAACCTTCATGTTTTCATCTGAGAGTAACTTCTTGT 3802
      |||||::: |||||:::
802 erAspIleGlyGlnGlyMetAspTyrIleValLysSerLeuSerSerGlu 818
      |||||::: |||||:::
3803 CGGATTTTGGGTCTCATGACTAGCTGTTGTTTC..... 3837
      |||||::: |||||:::
819 ProCys...AspLeuGlnGluIleGlnLeuValSerCysLeuSerAl 834
      |||||::: |||||:::
3838 ...TGTAAAGAACTCACAGAAATTAAGTTTTCGAGATTCATTTT... 3879
      |||||::: |||||:::
834 aAsnAlaValLysIleLeuAlaGlnAsnLeuHisAsnLeuValLysLeus 851
      |||||::: |||||:::
3880 .CAAGCGTCCCATTTGTTGCC...AGTTTGCCAAATTTATTTCTCTGA 3925
      |||||::: |||||:::

```



```

1      RESULT 2
2      US-08-569-749-13
3      : Sequence 13, Application US/08569749
4      : Patent No. 6187557
5      : GENERAL INFORMATION:
6      : APPLICANT: Rothe, Mike
7      : TITLE OF INVENTION: Goedel, David V
8      : NUMBER OF SEQUENCES: 14
9      : CORRESPONDENCE ADDRESS: 1
10     : ADDRESSEE: FLEHN, HOHBACH, TEST, ALBRITTON & HERBERT
11     : STREET: 4 Embarcadero Center, Suite 3400
12     : CITY: San Francisco
13     : STATE: California
14     : COUNTRY: USA
15     : ZIP: 94111
16     : COMPUTER READABLE FORM:
17     : MEDIUM TYPE: Floppy disk
18     : COMPUTER: IBM PC compatible
19     : OPERATING SYSTEM: PC-DOS/MS-DOS
20     : SOFTWARE: Patent in Release #1.0, Version #1.30
21     : CURRENT APPLICATION DATA:
22     : APPLICATION NUMBER: US/08/569,749
23     : FILING DATE:
24     : CLASSIFICATION: 514
25     : ATTORNEY/AGENT INFORMATION:
26     : NAME: Brezner, David J
27     : REGISTRATION NUMBER: 24,774
28     : REFERENCE/DOCKET NUMBER: A-62464/DJB
29     : TELECOMMUNICATION INFORMATION:
30     : TELEPHONE: (415)781-1989
31     : TELEFAX: (415)398-1249
32     : INFORMATION FOR SEQ ID NO: 13:
33     : SEQUENCE CHARACTERISTICS:
34     : LENGTH: 2862 base pairs
35     : TYPE: nucleic acid
36     : STRANDEDNESS: single
37     : TOPOLOGY: linear
38     : MOLECULE TYPE: cDNA
39     : US-08-569-749-13

```

Query Match	1.33;	Score 39.6;	DB 4;	Length 2862;
Best Local Similarity	46.98;	Pred. No. 0.16;		
Matches 123;	Conservative 0;	Mismatches 139;	Indels 0;	Gaps 0

QY	85	tgactgttataaagcgaatcaacgaagaccatttctatcgatgtctgaatcgagag	144
Db	1869	TGACACATGTCTCCATCTCGATATCTTGTAAGGCCAGTGAATTTACAAAACAGG	19228
QY	145	aagtaaacatcatttgcctcgagaaggtggaacagatgcctgctagaagatcaltaca	204
Db	1929	AMCATGTATTTATTATAGCAGAAAAACACATATCCCTTACAAACGAAGAGCTTATTGACA	19888
QY	205	tgatttgaataaagggttcaagagtcctgtgaacctcttcttaatcccttaagaggtga	264
Db	1989	CCGTTTATAGTCAAAGGAAATGCTGCGCCCAACATCTTCAAAAACCTCTGGAAGGAAATTG	20488
QY	265	actatccctcatcttcagagacttgaaaggaacaaagtcttctcatcaagatcaagaagag	324
Db	2049	ACTCCACGTTATATGAAAACTTATTGTGTGAAAAAGATATGAAAGTATTTCCACACAGAG	21088
QY	325	acttggacgaatttgctgcagga	346
Db	2109	ACGTTTCAGGCTTGTCATTGGA	2130

```

RESULT      3
PCT-US96-12860-13
; Sequence 13, Application PC/TUS9612860
; GENERAL INFORMATION:
; APPLICANT:  TULARIK, INC.

```

1 TITLE OF INVENTION: INHIBITORS OF APOPTOSIS
 2 NUMBER OF SEQUENCES: 14
 3 CORRESPONDENCE ADDRESS:
 4 ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
 5 STREET: 4 Embarcadero Center, Suite 3400
 6 CITY: San Francisco
 7 STATE: California
 8 COUNTRY: USA
 9 ZIP: 94111
 10
 11 COMPUTER READABLE FORM:
 12 MEDIUM TYPE: floppy disk
 13 COMPUTER: IBM PC compatible
 14 OPERATING SYSTEM: PC-DOS/MS-DOS
 15 SOFTWARE: Patent Release #1.0, Version #1.30
 16 CURRENT APPLICATION DATA:
 17 APPLICATION NUMBER: PCT/US96/12860
 18 FILING DATE: 06 AUG 1996
 19
 20 CLASSIFICATION:
 21 PRIOR APPLICATION DATA:
 22 APPLICATION NUMBER: U.S. Serial Nos. 08/512,946 & 08/569,749
 23 CLASSIFICATION:
 24 ATTORNEY/AGENT INFORMATION:
 25 NAME: Brezner, David J.
 26 REGISTRATION NUMBER: 24,774
 27 REFERENCE/DOCKET NUMBER: A-62464/DJB
 28 TELECOMMUNICATION INFORMATION:
 29 TELEPHONE: (415)781-1989
 30 TELEFAX: (415)398-3249
 31 INFORMATION FOR SEQ ID NO: 13:
 32 SEQUENCE CHARACTERISTICS:
 33 LENGTH: 2862 base pairs
 34 TYPE: nucleic acid
 35 STRANDEDNESS: single
 36 TOPOLOGY: linear
 37
 38 MOLECULE TYPE: CDNA
 39
 40 PCT-US96-12860-13

```
Query Match      1.3%; Score 39.6; DB 5; Length 2862;
Best Local Similarity 46.9%; Pred. No. 0.16;
Matches 123; Conservative 0; Mismatches 139; Indels 0; Gaps 0;
```

QY	85	tgaatgataaagcaaacacagatgaacctattgatgagatgctcgaatcgag	144
Db	1869	TGACACATGTCCTCCATCTCTGGATATCTTTTAGGCCACGTGAATTACAAAACAG	1922
QY	145	aagtaaacatcaattgcctgcgagaagctgagcagaatgctgataaggaatcatca	204
Db	1929	AACATGATATTTATTAGACAGAAAACACAGATACCTTCACACCAAGAGCTTATTGACA	1988
QY	205	tgaattgaaaaagggttcagagtcctgttaacctcttcttaaccccttaagagctga	264
Db	1989	CCGTTTATGTCAAAGGGAATATGCTGCACGCCAACATCTCTCAAAAACCTCTGAAGGAAATTG	2048
QY	265	actaacctcatcttcagagctgaaaggaacaaagctcttcttcatacagacataagaagag	324
Db	2049	ACGTCCAGCTTATATGAAACTATTTGTGTGGAAAAGATATGAAGTATATTCACACAGAG	2108
QY	325	acttgaagcaattgcgtctcaga	346
Db	2109	ACGTTTCAGGCTGTGTCATTGGA	2130

```

RESULT 4
US-09-212-971-13
; Sequence 13, Application US/09212971HE
; Patent No. 6107041
; GENERAL INFORMATION:
; APPLICANT: Korneluk, Robert G
; APPLICANT: MacKenzie, Alexander E
; APPLICANT: Liston, Peter
; APPLICANT: Baird, Stephen
; APPLICANT: Tsang, Benjamin K

```

APPLICANT: Pratt, Christine
TITLE OF INVENTION: DETECTION AND MODULATION OF IAPS AND
TITLE OF INVENTION: NAIP FOR THE DIAGNOSIS AND TREATMENT OF PROLIFERATIVE
TITLE OF INVENTION: DISEASE
FILE REFERENCE: 07891/009002
CURRENT APPLICATION NUMBER: US/09/212,971B
CURRENT FILING DATE: 1998-12-16
EARLIER APPLICATION NUMBER: 60/017,354
EARLIER FILING DATE: 1996-04-26
EARLIER APPLICATION NUMBER: 60/030,590
EARLIER FILING DATE: 1996-11-14
EARLIER APPLICATION NUMBER: 08/800,929
NUMBER OF SEQ ID NOS: 17
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO: 13
LENGTH: 3151
TYPE: DNA
ORGANISM: Mus musculus
US-09-212-971-13

Query Match 1.3%; Score 39.6; DB 3; Length 3151;
Best Local Similarity 46.9%; Pred. No. 0.17;
Matches 123; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

QY 85 tgactgtataaagcaatcacagatgactatttgatggaatgtctgaatcgcgaag 144
DB 2175 tgacacatgctctctctatctctgataatctcttgaggcagtgtaattacaacagg 2234
QY 145 aagtaaacatcatctgctgagaaagtgagcagagatgctgctagagggatcaca 204
DB 2235 aacatgatatattagcagaaacacagatacccttaacagcagagctattgaca 2294
QY 205 tgatttgaaaagggtcagagtcctgttaacctcttctaactcccttaaggagtga 264
DB 2295 ccgttttagtcaggaagatgctgcagccacatcttcaaaactctctgaaggaaatg 2354
QY 265 acatcctctatttcaggacttgaatgagacaagctcttctcagactcgaagagag 324
DB 2355 acctcaggtatatagaacttatttctgtaaaaagaatagatattatccacagaag 2414
QY 325 acttgagcagattggctcaga 346
DB 2415 acgtttcagcgtctgctatgga 2436

RESULT 5

US-08-800-929A-13
Sequence 13, Application US/08800929A
Patent No. 6133437
GENERAL INFORMATION:
APPLICANT: Korneluk, Robert G
APPLICANT: Mackenzie, Alexander E
APPLICANT: Liston, Peter
APPLICANT: Baird, Stephen
APPLICANT: Tsang, Benjamin K
APPLICANT: Pratt, Christine
TITLE OF INVENTION: DETECTION AND MODULATION OF
TITLE OF INVENTION: IAPS AND NAIP FOR THE DIAGNOSIS AND TREATMENT OF PROLIFERATIVE
TITLE OF INVENTION: DISEASE
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Clark & Elding LLP
STREET: 176 Federal Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/800,929A
FILING DATE: 13-FEB-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/030,590
FILING DATE: 14-NOV-1996
APPLICATION NUMBER: 60/017,354
FILING DATE: 26-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Bleker-Brady, Kristina
REGISTRATION NUMBER:
REFERENCE/DOCKET NUMBER: 07891/009001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-428-0200
TELEFAX: 617-428-7045
TELEX:
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 3151 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-800-929A-13

Query Match 1.3%; Score 39.6; DB 3; Length 3151;
Best Local Similarity 46.9%; Pred. No. 0.17;
Matches 123; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

QY 85 tgactgtataaagcaatcacagatgactatttgatggaatgtctgaatcgcgaag 144
DB 2175 tgacacatgctctctctatctctgataatctcttgaggcagtgtaattacaacagg 2234
QY 145 aagtaaacatcatctgctgagaaagtgagcagagatgctgctagagggatcaca 204
DB 2235 AACATGATATTATTAGACAGAAACACAGATACCCCTTACAAAGCAGAGACTTATTGACA 2294
QY 205 tgatttgaaaagggtcagagtcctgttaacctcttctaactcccttaaggagtga 264
DB 2295 CCGTTTATGCTCAAGGAATGCTCAGCCAAACATCTTCAAAACTCTCTGAGGAATTTG 2354
QY 265 acatcctctatttcaggacttgaatgagacaagctcttctcagactcgaagagag 324
DB 2355 ACTCCAGCTTATATGAACCTTATTGTGAAAAGAAATGAGATATATTCACACAGAG 2414
QY 325 acttgagcagattggctcaga 346
DB 2415 ACGTTTCAGGCTTGTCTATTGGA 2436

RESULT 6

US-09-617-053A-13
Sequence 13, Application US/09617053A
Patent No. 6300492
GENERAL INFORMATION:
APPLICANT: Korneluk, Robert G
APPLICANT: Mackenzie, Alexander E
APPLICANT: Liston, Peter
APPLICANT: Baird, Stephen
APPLICANT: Tsang, Benjamin K
APPLICANT: Pratt, Christine
TITLE OF INVENTION: DETECTION AND MODULATION OF IAPS AND
TITLE OF INVENTION: NAIP FOR THE DIAGNOSIS AND TREATMENT OF PROLIFERATIVE
TITLE OF INVENTION: DISEASE
FILE REFERENCE: 07891/009003
CURRENT APPLICATION NUMBER: US/09/617,053A
CURRENT FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: US 08/800,929
PRIOR FILING DATE: 1997-02-13
NUMBER OF SEQ ID NOS: 17

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 3151
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-617-053A-13

```

```

Query Match      1.3%; Score 39.6; DB 4; Length 3151;
Best Local Similarity 46.9%; Pred. No. 0.17;
Matches 123; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

```

```

QY 85 tgacgtatataagaacatcacagatgaccttttgtatgaatgtctgaatcgcaag 144
DB 2175 tgaacatgcttcctcctcctcctgaataactcttgaggccagtgtaattacaacagg 2234
      145 aaglaaacatcattgctgcgagaagtgtagcagatgctctgaaggaggaattcaca 204
      2235 aacatgatatatatagacagaaacacacagatcccttaacaagcaagaggtattgaca 2294
QY 205 tgatttgaaaaaagggttcagagtcctgtaacctcttcttaaccttaaggagtggga 264
DB 2295 ccgttttagcaaggagaatgctgcagccaacatcttcaaaaacctctgaaggaaattg 2354
QY 265 aactaccctcttttcagagacttgatgacaaagctcttttcatcagaatcagaagag 324
DB 2335 actccaagctataatgaactatttggaaaagaataatgatataltccaacagaag 2414
QY 325 acttgacgatttggtcctcaga 346
DB 2415 acgttcaggctgtgcatggga 2436

```

RESULT 7

```

; Sequence 1, Application PC/TUS9505922A
; GENERAL INFORMATION:
; APPLICANT: HE, ET. AL.

```

```

; TITLE OF INVENTION: Human Inhibitor of Apoptosis Gene 1
; NUMBER OF SEQUENCES: 8

```

```

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,

```

```

; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND

```

```

; STATE: NEW JERSEY
; COUNTRY: USA

```

```

; ZIP: 07068
; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2

```

```

; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/05922A

```

```

; FILING DATE: 11 MAY 1995
; CLASSIFICATION:

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:

```

```

; FILING DATE:
; ATTORNEY/AGENT INFORMATION:

```

```

; NAME: FERRARO, GREGORY D.
; REGISTRATION NUMBER: 36,134

```

```

; REFERENCE/DOCKET NUMBER: 325800-292
; TELECOMMUNICATION INFORMATION:

```

```

; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744

```

```

; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 1435 BASE PAIRS
; TYPE: NUCLEIC ACID

```

```

; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR

```

```

; MOLECULE TYPE: CDNA
PCT-US95-05922A-1

```

```

Query Match      1.2%; Score 36.8; DB 5; Length 1435;
Best Local Similarity 46.2%; Pred. No. 0.71;
Matches 122; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

```

```

QY 83 aatgactgtataaagaacatcacagatgacctatttgtatgaatgtctgaatcgca 142
DB 882 ATTGACATGTGTCCTTCCTCATCTCGATATATCTTTTAAAGCCCATGTATTAATAACA 941
QY 143 agaagtaaacatcaatattgtctgcgagaagtgtagcagaatgtcgtctagaaggatcattca 202
DB 942 GGAACATGATATTATTAATAACAAAAACACAGATACCTTTACAAAGCAGAGAACTGATTGA 1001
QY 203 catgtatttggaaaaagggttcagagtcctgtaacctcttcttaaccttaaccttaaggagtg 262
DB 1002 TACCATTTTGGTTAAAGAAATGCTGCGCAACATCTTCAAAAACCTGTAAAGAAAT 1061
QY 263 gaactaccctcatcttcagagacttgatgacaaagctcttctcatcagacatcagaag 322
DB 1062 TGACTCTACATTTGTATTAACAACTTATTGTGATAGATATGAGTATATTCACACAGA 1121
QY 323 agacttgacagatttggtcctcaga 346
DB 1122 AGATGTTTCAGCTCTGTCACTGGA 1145

```

RESULT 8

```

; Sequence 1, Application US/08569749
; Patent No. 6187557

```

```

; GENERAL INFORMATION:
; APPLICANT: Rothe, Mike

```

```

; TITLE OF INVENTION: INHIBITORS OF APOPTOSIS
; NUMBER OF SEQUENCES: 14

```

```

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT

```

```

; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco

```

```

; STATE: California
; COUNTRY: USA

```

```

; ZIP: 94111
; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/569,749

```

```

; FILING DATE:
; CLASSIFICATION: 514

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Brezner, David J.

```

```

; REGISTRATION NUMBER: 24,774
; REFERENCE/DOCKET NUMBER: A-62464/DJB

```

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)781-1989

```

```

; TELEFAX: (415)396-3249
; INFORMATION FOR SEQ ID NO: 1:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 2589 base pairs

```

```

; TYPE: nucleic acid
; STRANDEDNESS: single

```

```

; TOPOLOGY: linear
; MOLECULE TYPE: CDNA

```

```

US-08-569-749-1

```

```

Query Match      1.2%; Score 36.8; DB 4; Length 2589;
Best Local Similarity 46.2%; Pred. No. 1;

```


; SEQUENCE CHARAC

TELEX: 25-3856

QY 1561 tcgcagcagtgatcaacaacgcgtgctcttcggaacttccatcgccaagagcctctc 1620
|||||
Db 1561 tcgcagcagtgatcaacaacgcgtgctcttcggaacttccatcgccaagagcctctc 1620
QY 1621 ggagacaggaactcttgcgaagtgtgaaaacaccactgagcaagaaattctgaaagcca 1680
|||||
Db 1621 ggagacaggaactcttgcgaagtgtgaaaacaccactgagcaagaaattctgaaagcca 1680
QY 1681 taacatcatatctcttctgtgagagtggtgcatttatatacaagagatcatccaaat 1740
|||||
Db 1681 taacatcatatctcttctgtgagagtggtgcatttatatacaagagatcatccaaat 1740
QY 1741 cagcccccagacccaagaattgaaagcttctcttcaaggtaaaagcttatatacaactcag 1800
|||||
Db 1741 cagcccccagacccaagaattgaaagcttctcttcaaggtaaaagcttatatacaactcag 1800
QY 1801 ggaacatccccgaattacttacttcttgaactcttgaacatttgcgaattgtgcaagtgc 1860
|||||
Db 1801 ggaacatccccgaattacttacttcttgaactcttgaacatttgcgaattgtgcaagtgc 1860
QY 1861 tggacttcatataactggaacttcttcttcttcttcttcttcttcttcttcttcttctt 1920
|||||
Db 1861 tggacttcatataactggaacttcttcttcttcttcttcttcttcttcttcttcttctt 1920
QY 1921 aagacacaggtggaatccacaatggaagagccccaagaactatactccagcagaggtg 1980
|||||
Db 1921 aagacacaggtggaatccacaatggaagagccccaagaactatactccagcagaggtg 1980
QY 1981 taatcttcttcttcaacttgaagcaggaattcaagactctggaagtgcacactccggagt 2040
|||||
Db 1981 taatcttcttcttcaacttgaagcaggaattcaagactctggaagtgcacactccggagt 2040
QY 2041 tcagcaagtggaataagaataatatacatatctctggaagaaatattcaagctctccaa 2100
|||||
Db 2041 tcagcaagtggaataagaataatatacatatctctggaagaaatattcaagctctccaa 2100
QY 2101 gccctcaggtcgaataaagaagatgtgctgtgtgctggaagcctcaggttgtgtctca 2160
|||||
Db 2101 gccctcaggtcgaataaagaagatgtgctgtgtgctggaagcctcaggttgtgtctca 2160
QY 2161 gcaacctgtaagaacattatctctcctcatgtgtggaagccagtcctccatcagataga 2220
|||||
Db 2161 gcaacctgtaagaacattatctctcctcatgtgtggaagccagtcctccatcagataga 2220
QY 2221 agagcgacatcatatctctcctcatgtgtggaagccagtcctccatcagataga 2280
|||||
Db 2221 agagcgacatcatatctctcctcatgtgtggaagccagtcctccatcagataga 2280
QY 2281 aacggctgcgggtgtgtctgctgacagcttgggttaacttgaagaaccttacaagctca 2340
|||||
Db 2281 aacggctgcgggtgtgtctgctgacagcttgggttaacttgaagaaccttacaagctca 2340
QY 2341 taatggaatacaataaagaatgaagaagaatgctataaatactagctgaagcctgaa 2400
|||||
Db 2341 taatggaatacaataaagaatgaagaagaatgctataaatactagctgaagcctgaa 2400
QY 2401 accgtgaagaagaatgtgttatttatttatttatttatttatttatttatttatttatt 2460
|||||
Db 2401 accgtgaagaagaatgtgttatttatttatttatttatttatttatttatttatttatt 2460
QY 2461 attaatatgataatctctctgcaagtgaaacctgtgacctggaagaataatcaatagct 2520
|||||
Db 2461 attaatatgataatctctctgcaagtgaaacctgtgacctggaagaataatcaatagct 2520
QY 2521 cctgctgtctgtctgcaaatgtcagtgaaacctgtgacctggaagaataatcaatagct 2580
|||||
Db 2521 cctgctgtctgtctgcaaatgtcagtgaaacctgtgacctggaagaataatcaatagct 2580
QY 2581 aacgtgagatctcttgattatcaagaataatactcctggaagaagaatgaagaatctctc 2640
|||||
Db 2581 aacgtgagatctcttgattatcaagaataatactcctggaagaagaatgaagaatctctc 2640
QY 2641 atgacatgatcagacagatgtagcgtgtagaacagctcacccagctgtagtgccttgg 2700

Db 2641 atgacatgatcagacagatgtagcgtgtagaacagctcacccagctgtagtgccttgg 2700
|||||
QY 2701 gctgtgacgtgcgaagcagcctgagcagcctgttgaacatttgaagaggtcccaaac 2760
|||||
Db 2701 gctgtgacgtgcgaagcagcctgagcagcctgttgaacatttgaagaggtcccaaac 2760
QY 2761 tcgctcaagcttgggttgaagaacctggaagactcacagatcacagagatagaaatttagtg 2820
|||||
Db 2761 tcgctcaagcttgggttgaagaacctggaagactcacagatcacagagatagaaatttagtg 2820
QY 2821 catttttgaagaagaaccttgaagaacctccagcagcttgaacttgcgggaatcgtg 2880
|||||
Db 2821 catttttgaagaagaaccttgaagaacctccagcagcttgaacttgcgggaatcgtg 2880
QY 2881 tgaagcagtgatgtagtgccttctcctcatggtgtatattgagaatcttaagaattagtg 2940
|||||
Db 2881 tgaagcagtgatgtagtgccttctcctcatggtgtatattgagaatcttaagaattagtg 2940
QY 2941 ttttgaacttgaacttaagaatttctacactgataccagcattgtagaagaattagcc 3000
|||||
Db 2941 ttttgaacttgaacttaagaatttctacactgataccagcattgtagaagaattagcc 3000
QY 3001 aaggttataccaagttaacttcttctgcaagaagctagctgttgggtggaattgag 3060
|||||
Db 3001 aaggttataccaagttaacttcttctgcaagaagctagctgttgggtggaattgag 3060
QY 3061 atgatgatctcagtgattatcaagtgctttaaactagtaactcgtcttaataaagtgt 3120
|||||
Db 3061 atgatgatctcagtgattatcaagtgctttaaactagtaactcgtcttaataaagtgt 3120
QY 3121 ctggaagcagta 3133
|||||
Db 3121 ctggaagcagta 3133
RESULT 2
AAH98254
ID AAH98254 standard; cDNA: 3545 BP.
XX
AC AAH98254;
XX
DT 12-OCT-2001 (first entry)
XX
DE Murine EST-derived coding sequence SEQ ID NO: 111.
XX
KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder;
KW biodiversity; gene therapy; nutrition; ss.
XX
OS Mus musculus.
XX
PN W0200154477-A2.
XX
PD 02-AUG-2001.
XX
FE 25-JAN-2001; 2001WO-US02687.
XX
PR 25-JAN-2000; 2000US-0491404.
PR 17-JUL-2000; 2000US-061746.
PR 03-AUG-2000; 2000US-0631451.
PR 15-SEP-2000; 2000US-0663870.
XX
PA (HXSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI Cao Y, Drmanac RA, Zhang J, Werhman T;
XX
DR WPI: 2001-476164/51.
DR P-PSDB: AAM23595.
XX
PT Isolated polypeptide for treatment of diseases, diagnostics, raising

Dh	2058	ggaattcaataactggtgccttattatgvggagctatgctcaatgvggaaagctgcaga	2117
Oy	1922	agacacaggtgtaataccacaatggaagagggccacagaacactcaatccacagagctgt	1981
Dh	2118	agacacaggtgtaataccacaatggaagagggccacagaacactcaatccacagagctgt	2177
Oy	1982	atcttgttcttcaactcgtgaagcaggaatctcagacctcgtgaggtcacactccggat	2041
Dh	2178	atcttgttcttcaactcgtgaagcaggaatctcagacctcgtgaggtcacactccggat	2237
Oy	2042	cagcaaatgttaataagaagaataccaatactcgggaaaataattcaactctgcacaga	2101
Dh	2238	cagcaaatgttaataagaagaataccaatactcgggaaaataattcaactctgcacaga	2297
Oy	2102	cccaagagctgcgaataaagaagatgtcagctgtgtgtgtgtgaagccctcaattgtgtccag	2161
Dh	2298	cccaagagctgcgaataaagaagatgtcagctgtgtgtgtgtgaagccctcaattgtgtccag	2357
Oy	2162	caactgtlaagaaactatctctctcatgtgtggaagccagtcacctacacataagaatga	2221
Dh	2358	caactgtlaagaaactatctctctcatgtgtggaagccagtcacctacacataagaatga	2417
Oy	2222	gagggacaatacaactcgtttaaacaaccttgaaaacctgtgattctacatgacctacagaatca	2281
Dh	2418	gagggacaatacaactcgtttaaacaaccttgaaaacctgtgattctacatgacctacagaatca	2477
Oy	2282	acggctcccggtgtgtctgtgactacagctgtgtgttaacttgagaaccttacaagaactcat	2341
Dh	2478	acggctcccggtgtgtctgtgactacagctgtgtgttaacttgagaaccttacaagaactcat	2537
Oy	2342	aaatgatacaataaagaatgaatgaagaagaatgtctaataaactagacttgaagcctgnaaaa	2401
Dh	2538	aaatgatacaataaagaatgaatgaagaagaatgtctaataaactagacttgaagcctgnaaaa	2597
Oy	2402	ccgtgaagaagaatgttattattcaattggccaactgtctgtacatcttgagaaggagaaatga	2461
Dh	2598	ccgtgaagaagaatgttattattcaattggccaactgtctgtacatcttgagaaggagaaatga	2637
Oy	2462	ttacaatgtcaagacctctgtgtcaagtgaaacctgtgaccttgagaagaattcaatagtctc	2521
Dh	2658	ttacaatgtcaagacctctgtgtcaagtgaaacctgtgaccttgagaagaattcaatagtctc	2717
Oy	2522	ctgtgctgtgtctgtcaaatgcaatgcaaatcctagctcagaatctcttcaaatctgtgtcaa	2581
Dh	2718	ctgtgctgtgtctgtcaaatgcaatgcaaatcctagctcagaatctcttcaaatctgtgtcaa	2777
Oy	2582	actgaagcaatcttgattatccagaanaatactactcgtgaanaaagtgtgaatgaagctcttca	2641
Dh	2778	actgaagcaatcttgattatccagaanaatactactcgtgaanaaagtgtgaatgaagctcttca	2837
Oy	2642	tgaactatcgaacagagatgaacgtgtgtagaagaagctccacagcaactgagatgagcccttggg	2701
Dh	2838	tgaactatcgaacagagatgaacgtgtgtagaagaagctccacagcaactgagatgagcccttggg	2897
Oy	2702	ctgtgtgacgtgtgcaagcagccttgagcagcctgtgtgaanaatttgagaagagttccacaact	2761
Dh	2898	ctgtgtgacgtgtgcaagcagccttgagcagcctgtgtgaanaatttgagaagagttccacaact	2957
Oy	2762	cgtaacagctgtgtgtgaanaaactcgtgagactccaagaatacagaagatttaaatctttaggtgc	2821
Dh	2958	cgtaacagctgtgtgtgaanaaactcgtgagactccaagaatacagaagatttaaatctttaggtgc	3017
Oy	2832	attctttggaagaagacccctctgnaaaacttccacagcagttggaatttgggcgggaatactgtt	2881
Dh	3018	attctttggaagaagacccctctgnaaaacttccacagcagttggaatttgggcgggaatactgtt	3077
Oy	2882	gagcagatgaatgaatgcttgccttcaatgggtgattttggaatcttcaagcaattagtggt	2941
Dh	3078	gagcagatgaatgaatgcttgccttcaatgggtgattttggaatcttcaagcaattagtggt	3137
Oy	2942	ttttggaacttaagtactaaagaattcttcaactgataccagcattagtctcagaanaacttagcca	3001
Dh	3138	ttttggaacttaagtactaaagaattcttcaactgataccagcattagtctcagaanaacttagcca	3197

QY	3002	agcgtatcccaagtaactcttctcgcaagaagcctaggctgttgggtggcaattgtaga	3061
Db	3198	agtgtatcccaagtaactcttctcgcaagaagcctaggctgttgggtggcaattgtaga	3257
QY	3062	tgatcatcctcagttctttatcaggtgccttttaactcgtactcgtcttaataagctgac	3121
Db	3258	tgatcatcctcagttctttatcaggtgccttttaactcgtactcgtcttaataagctgac	3317
QY	3122	tcgaagccagta	3133
Db	3318	tcgaagccagta	3329
RESULT	3		
ID	AA503946	standard; DNA; 3615 BP.	
XX	AA503946;		
AC	12-SEP-2001	(first entry)	
DT			
DE	Human caspase recruitment domain 12 (CARD-12) genomic DNA.		
XX			
KM	Caspase recruitment domain; CARD-12; apoptosis; stress-related pathway;		
KM	cancer; viral infection; poxvirus; adenovirus; autoimmune disorder;		
KM	systemic lupus erythematosus; arthritis; neurological disorder; stroke;		
KM	Alzheimer's disease; amyotrophic lateral sclerosis; hematologic disease		
KM	aplastic anaemia; myocardial infarction; inflammatory disorder;		
KM	Crohn's disease; insulin-dependent diabetes; contact dermatitis;		
KM	psoriasis; graft rejection; bacterial infection; lepromatous leprosy;		
KM	tuberculosis; ischemic brain injury; hypoxic brain injury; ds;		
KM	kidney ischemia; reperfusion injury; acute bacterial meningitis;		
KM	excitotoxic brain damage; liver disease.		
OS	Homo sapiens.		
FH	Key	Location/Qualifiers	
FT	CDS	1..3615	
FT		/*tag= a	
FT		/product= "Human CARD-12"	
XX	WO200130971-A2.		
XX	03-MAY-2001.		
PD			
XX	26-OCT-2000; 2000WO-US29643.		
PF			
XX	27-OCT-1999; 99US-0161822.		
PR			
XX	(MILL-) MILLENNIUM PHARM INC.		
PA			
XX	Bertin J, Robison KE;		
PI	WPI; 2001-308628/32.		
DR	P-PSDB; AAU02881.		
XX			
PT	Isolated caspase recruitment domain-12 polypeptide and nucleic acids		
PT	encoding them, useful for treating and diagnosing disorders associated		
PT	with abnormal apoptosis such as cancer, arthritis and Alzheimer's		
PT	disease -		
XX			
PS	Disclosure; Fig 2; 93pp; English.		
XX			
XX	The sequence represents a genomic DNA which encodes the human caspase		
CC	recruitment domain 12 (CARD-12) polypeptide. CARD domains are found in a		
CC	number of proteins that transmit signals that activate apoptosis and		
CC	inflammatory pathways in response to stress and other stimuli. Therefore		
CC	CARD-12 and its corresponding nucleic acid may be used in treatment and		
CC	diagnosis of patients suffering from disorders associated with an		
CC	abnormal level (an increase or a decrease) of apoptotic cell death or		
CC	abnormal activity of stress-related pathways. The disorders include		
CC	cancer, viral infections (e.g. caused by poxviruses, adenoviruses),		

CC autoimmune disorders (e.g. systemic lupus erythematosus, arthritis),
CC neurological disorders (e.g. Alzheimer's disease, amyotrophic lateral
CC sclerosis), haematologic diseases (e.g. aplastic anaemia, myocardial
CC infarction, stroke), inflammatory and immune system disorders (e.g.
CC Crohn's disease, insulin-dependent diabetes, contact dermatitis,
CC psoriasis, graft rejection), bacterial infections (e.g. tuberculosis,
CC dermatomycosis leprosy), ischaemic and hypoxic brain injury, kidney
CC ischaemia/reperfusion injury, excitotoxic brain damage, acute bacterial
CC meningitis and liver disease.

SQ Sequence 3615 BP; 1041 A; 811 C; 845 G; 918 T; 0 other;

Query Match	95.1%	Score 2980.6;	DB 22;	Length 3615;
Best Local Similarity	98.1%	Pred. NO. 0;		
Matches 3047; Conservative	0;	Mismatches	4;	Indels 54; Gaps 1

QY	37	tgaatttcataaagagaaataagcccgagcccttatccaagaatbtygaatgactygtataa	96
	419	tgaatttcataaagaaataagcaataagcccgagcccttatccaagaatbtygaatgactygtataa	478
	97	agcaaatcacagaatgagacctattgtatbtygaatgttcctgaatcgcgaaagatcaaacatca	156
Db	479	agcaaatcacagaatgagacctattgtatbtygaatgttcctgaatcgcgaaagatcaaacatca	538
QY	157	tttgcgcgagaagctgagacagatgctgtctagagagatcatcatcaatgtatttgaana	216
Db	539	tttgcgcgagaagctgagacagatgctgtctagagagatcatcatcaatgtatttgaana	598
QY	217	agggttcaagagctcctgtgaacctctctcttaaacctcttaagagatgagaaactctctcat	276
Db	599	agggttcaagagctcctgtgaacctctctcttaaacctcttaagagatgagaaactctctcat	658
QY	277	ttcagagacttgaatbtygaacaaa-----	297
Db	659	ttcagagacttgaatbtygaacaaagtttbaggagacacagaaatbtygctctcttaaacatca	718
QY	298	-----gcttttttcatacagacatacgaagaagagaaacttgcagaaatttgcctc	342
Db	719	cctctctctaataagctctcttttcatacagacatacgaagaagagaaacttgcagaaatttgcctc	778
QY	343	agaaattcaagagactgtaccatacccaactctttctgaactttaaccctctgttgaag	402
Db	779	agaaattcaagagactgtaccatacccaactctttctgaactttaaccctctgttgaag	838
QY	403	ataatgacataattttaacttgcanaaagcaaccttcacagaaactgtcctgttgcgagaag	462
Db	839	ataatgacataattttaacttgcanaaagcaaccttcacagaaactgtcctgttgcgagaag	898
	463	accaaaaccatacccgctgagacgaactgaaccttgaaatgctcctctgcagagctcttcaga	522
Db	899	accaaaaccatacccgctgagacgaactgaaccttgaaatgctcctctgcagagagctcttcaga	958
QY	523	gcccctgcacatctgaaaggggaaatctgcgcgaagaagcaagccaactctgcgcgcgacatg	582
Db	959	gcccctgcacatctgaaaggggaaatctgcgcgaagaagcaagccaactctgcgcgcgacatg	1018
QY	583	ccatgctctcgggctcccggaagctgcgaagctctgcaccaagtccaatattgctctcttc	642
Db	1019	ccatgctctcgggctcccggaagctgcgaagctctgcaccaagtccaatattgctctcttc	1078
QY	643	tcgcgtcccaagaagcccaaggctgagaaacttttgaaacccctctgtgataaacctcgata	702
Db	1079	tcgcgtcccaagaagcccaaggctgagaaacttttgaaacccctctgtgataaacctcgata	1138
QY	703	taactgcacaaatcagaagaacagacactcatatgcgcacatgctgtctgcgaagctgcgcgcgaag	762
Db	1139	taactgcacaaatcagaagaacagacactcatatgcgcacatgctgtctgcgaagctgcgcgcgaag	1198
QY	763	ttctttctctcttgaatgctacatgaatccaagccccaagaactgcgccaagaatcagaag	822
	1199	ttctttctctcttgaatgctacatgaatccaagccccaagaactgcgccaagaatcagaag	1258

QY	823	ccctgataaagaaacaccacgccttcaagaacatggtcactgtaaccacttaccactgagt	882
Db	1259	ccctgtataaagaaacaccacgccttcaagaacatggtcactgtaaccacttaccactgagt	1318
QY	883	gscctgagacataatcgccgagtttggtgccccttgactctgtaggtggtgggtatgacaagag	942
Db	1319	gscctgagacataatcgccgagtttggtgccccttgactctgtaggtggtgggtatgacaagag	1378
QY	943	aacagagcccaagcctcatalcccgagaagtgctgtatcaagagagcctgtgtcaaggtctgtgc	1002
Db	1379	aacagagcccaagcctcatalcccgagaagtgctgtatcaagagagcctgtgtcaaggtctgtgc	1438
QY	1003	tccaatctcagaataccaagtgctcttgagaatctatgaaagacccctctcttgtgtca	1063
Db	1439	tccaatctcagaataccaagtgctcttgagaatctatgaaagacccctctcttgtgtca	1498
QY	1063	tcaacttgacataccagatggtgtgaaagtgafttccaacttcaacacaaacaaacgcgt	1122
Db	1499	tcaacttgacataccagatggtgtgaaagtgafttccaacttcaacacaaacaaacgcgt	1586
QY	1123	tccaataccttctatgactgtgtgtatagaanaaacaaacaaacaaataaagtgctgtgc	1182
Db	1559	tccaataccttctatgactgtgtgtatagaanaaacaaacaaacaaataaagtgctgtgc	1618
QY	1183	caagtacttcaatctcgagaccctcgacactgtgtgagacctgactctgtgaggtgtgtct	1242
Db	1619	caagtacttcaatctcgagaccctcgacactgtgtgagacctgactctgtgaggtgtgtct	1678
QY	1243	ccccaagtttgatttggaactctgcagatgtgtccacggtgtgatatgagatgtgcctgcga	1302
Db	1679	ccccaagtttgatttggaactctgcagatgtgtccacggtgtgatatgagatgtgcctgcga	1738
QY	1303	caactgagcctcctctgtcaaatatatacaactcaaaaggttccaagccaaagatlaaatctctc	1362
Db	1739	caactgagcctcctctgtcaaatatatacaactcaaaaggttccaagccaaagatlaaatctctc	1798
QY	1363	acaagtaacttccagagatlacacagcagagcaagactcagacttatatgacgtctcaatg	1422
Db	1799	acaagtaacttccagagatlacacagcagagcaagactcagacttatatgacgtctcaatg	1858
QY	1423	agccagagagaggtgacccaaggggaatggttactcttgagaanaatggtttccattcttgaca	1488
Db	1859	agccagagagaggtgacccaaggggaatggttactcttgagaanaatggtttccattcttgaca	1918
QY	1483	ttacatccaactatagacagcctgtctccggtacacccgtgtgggtcatctgtggaagcacaca	1542
Db	1919	ttacatccaactatagacagcctgtctccggtacacccgtgtgggtcatctgtggaagcacaca	1978
QY	1543	gggcgtgtatgaaagcacctcgcagcagtgatatcaacagcgtgtgcttctcgtgacttcca	1602
Db	1979	gggcgtgtatgaaagcacctcgcagcagtgatatcaacagcgtgtgcttctcgtgacttcca	2038
QY	1603	tcgcgaagagcctctcttgtagagacaggaatctttgcaaaagtgtgaaaaacacactgagc	1663
Db	2039	tcgcgaagagcctctcttgtagagacaggaatctttgcaaaagtgtgaaaaacacactgagc	2098
QY	1663	aagaaattctgaagagcatalaacaatcaattcctttgaaagtgtgacatccatataatc	1722
Db	2099	aagaaattctgaagagcatalaacaatcaattcctttgaaagtgtgacatccatataatc	2158
QY	1723	aagagatgatacccaaatcagccctcgagcccaagaatlttgaagcttcttccaagtataaa	1782
Db	2159	aagagatgatacccaaatcagccctcgagcccaagaatlttgaagcttcttccaagtataaa	2218
QY	1783	gcttatataataactcaggggaacatcccgataacttatattgactctttgaaattgac	1842
Db	2219	gcttatataataactcaggggaacatcccgataacttatattgactctttgaaattgac	2278
QY	1843	gcaattgtgcaagtgcccttgactcttaataacttgaaacttttaigtggggaagctatgagtt	1902
Db	2279	gcaattgtgcaagtgcccttgactcttaataacttgaaacttttaigtggggaagctatgagctt	2338
QY	1903	catggtgaaaagcgtgcagagagacacaggttgaaattccaacttggaagagccccaagaacct	1962

Db 2339 catgggaaagcgtgcagaagacacagctggaatccacatgcaggaagggccccaagaacct 2398
QY 1963 acattcccgcaaggtctgtatcttcttcttcaactcggaagcggaatttaagaccttgg 2022
Db 2399 acattcccgcaaggtctgtatcttcttcttcaactcggaagcggaatttaagaccttgg 2458
QY 2023 aggtcaacacccgggatttcaagaattgataagaacatcacatatctggggaaaa 2082
Db 2459 aggtcaacacccgggatttcaagaattgataagaacatcacatatctggggaaaa 2518
QY 2083 tattcagctctgcacaaagcctcaggtctgcaaa taaagagatgtgctgtgtgtcgtgaa 2142
Db 2519 tattcagctctgcacaaagcctcaggtctgcaaa taaagagatgtgctgtgtgtcgtgaa 2578
QY 2143 gctcagcttctgtcctcagcagcctgttaagaaccttattctccatggtgtgaaagccaagtc 2202
Db 2579 gctcagcttctgtcctcagcagcctgttaagaaccttattctccatggtgtgaaagccaagtc 2638
QY 2203 cctccacatagaagattgaaggacacatcacatctgtaacaaacctgaaaccttgagta 2262
Db 2639 cctccacatagaagattgaaggacacatcacatctgtaacaaacctgaaaccttgagta 2698
QY 2263 ttcatgacctacagaa tcaacggctgcgggtgtgtcgtcagctgacagcttgggtgaacttga 2322
Db 2699 ttcatgacctacagaa tcaacggctgcgggtgtgtcgtcagctgacagcttgggtgaacttga 2758
QY 2323 agaacctacaaagctcataatgagtaacataaagaatgaatgaagaagatgctataaac 2382
Db 2759 agaacctacaaagctcataatgagtaacataaagaatgaatgaagaagatgctataaac 2818
QY 2383 tagctgaagcgctgtaaaacctgaagaagattgcttatttcaacttgaccacttcttg 2442
Db 2819 tagctgaagcgctgtaaaacctgaagaagattgcttatttcaacttgaccacttcttg 2878
QY 2443 acattgagaggggaatggaatgatacatgacagctcgtcgaag tgaacctgtgaccttg 2502
Db 2879 acattgagaggggaatggaatgatacatgacagctcgtcgaag tgaacctgtgaccttg 2938
QY 2503 aagaaattcaatagtctctcgtctgtctgtctgtcgaatgcaagtgaaacctgtctcaga 2562
Db 2939 aagaaattcaatagtctctcgtctgtctgtctgtcgaatgcaagtgaaacctgtctcaga 2998
QY 2563 atcttacaattgtgtcaaaactgagcattcttgattatagaagaaatttaccttgaaag 2622
Db 2999 atcttacaattgtgtcaaaactgagcattcttgattatagaagaaatttaccttgaaag 3058
QY 2623 atggaatgaagctctctcagactgacacagagatgaagctgctcagaaacagctcacg 2682
Db 3059 atggaatgaagctctctcagactgacacagagatgaagctgctcagaaacagctcacg 3118
QY 2683 caactgagtctgcctcgtgggtgtgtgacgttgcgaagcagcctgagcagcctgttgaacatt 2742
Db 3119 caactgagtctgcctcgtgggtgtgtgacgttgcgaagcagcctgagcagcctgttgaacatt 3178
QY 2743 tggaggaaggtcccaacactgctcaagcttgggttgaagaaacctggaacttcacagatcacg 2802
Db 3179 tggaggaaggtcccaacactgctcaagcttgggttgaagaaacctggaacttcacagatcacg 3238
QY 2803 aggttgaattttaggtgtcatttttggaaagaacctctgaaaaaccttccacagcttga 2862
Db 3239 aggttgaattttaggtgtcatttttggaaagaacctctgaaaaaccttccacagcttga 3298
QY 2863 atttgcgggaaatcgctgtgagcagtgatgagatgcttgccttcacatggtgtatttgaga 2922
Db 3299 atttgcgggaaatcgctgtgagcagtgatgagatgcttgccttcacatggtgtatttgaga 3358
QY 2923 atttgaagaattagtggtttttgactttagtaactaaagaatttccacctgacagcat 2982
Db 3359 atttgaagaattagtggtttttgactttagtaactaaagaatttccacctgacagcat 3418
QY 2983 tagtcaagaacttagcgaagtgtatccaagttaacttttcgcagaagaagctagcttg 3042

Db 3419 tagtcaagaacttagcgaagtgtatccaagttaacttttcgcagaagaagctagcttg 3478
QY 3043 ttgggtggcaatttgatgatgatgatcctcagtttatttaagctg 3087
Db 3479 ttgggtggcaatttgatgatgatgatcctcagtttatttaagctg 3523

RESULT 4
ID AAH99581
AAH99581 standard; cDNA, 2950 BP.
XX
AC AAH99581;
XX
DT 16-OCT-2001 (first entry)
XX
DE Human protein encoding cDNA sequence SEQ ID NO:416.
XX
KW Human; cancer; HIV infection; human immunodeficiency virus;
KW anti-inflammatory; antirheumatic; antiarthritic; immunosuppressive;
KW antibacterial; endocrine; cardiant; central nervous system; virucide;
KW anti-HIV; fungicide; antimitagen; cardiovascular; antinaemic; anaemia;
KW antiagregant; haemostatic; vulnerary; antiulcer; osteopathic; cytostatic;
KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
KW allergic rhinitis; diabetes; multiple sclerosis; depression;
KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
KW neurological disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200153455-A2.
XX
PD 26-JUL-2001.
XX
PF 22-DEC-2000; 2000MO-US35017.
XX
PR 23-DEC-1999; 99US-0471275.
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
PI
XX
DR WPI: 2001-457603/49.
XX
PT P-PSDB: AAM25640.
XX
PT Isolated human polynucleotides encoding polypeptides, useful for the
PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
XX
PS Claim 1: Page 511-512; 1217pp; English.
XX
CC AAH99166 to AAH9904 encode the human proteins given in AAM25225 to
CC AAM25963. The proteins can have activities based on the tissues and
CC cells they are expressed in, such as: antiinflammatory; antirheumatic;
CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;
CC central nervous system; virucide; anti-HIV; fungicide; antimitagen;
CC cardiovascular; antinaemic; antiagregant; haemostatic; vulnerary;
CC antiulcer; osteopathic; dermatological; antiallergic; antiasthmatic;
CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
CC encoding them can be used in gene therapy, antisense therapy and vaccine
CC production. The proteins and polynucleotides are useful for screening for
CC agonists or antagonists of a protein and for the treatment and diagnosis
CC of disorders associated with the activity of a protein e.g. inflammation,
CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC infections, autoimmunity, genetic diseases, haematopoietic disorders,

CC anaemia, platelet disorders, thrombocytopenia, wounds, burns, ulcers,
CC osteoporosis, severe combined immunodeficiency, eczema, allergic
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC neurological disorders.

XX Sequence 2950 BP; 915 A; 592 C; 628 G; 815 T; 0 other.

Query Match 62.1%; Score 1946.6; DB 22; Length 2950;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1949; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
OY 1181 tgcagtgactcattcgcgagccttgagaccactgtgagagccttaagctctgagagtggtt 1240
Db 1 tgcagtgactcattcgcgagccttgagaccactgtgagacattatctctcgagagtggtt 60

OY 1241 ctcccaagaatttgattctcgacactgcaagatgtgtccagcgttgatgagatgctcct 1300
Db 61 ctcccaagaatttgattctcgacactgcaagatgtgtccagcgttgatgagatgctcct 120

OY 1301 gacaactggtcctcctgtaataatacagctcaaaaggttcaagccaaagtataattctt 1360
Db 121 gacaactggtcctcctgtaataatacagctcaaaaggttcaagccaaagtataattctt 180

OY 1361 tcaagaatcattccagagagtaacaagcagagcagaagactcagcagttattgacgtctca 1420
Db 181 tcaagaatcattccagagagtaacaagcagagcagaagactcagcagttattgacgtctca 240

OY 1421 tgaagcagaagaggtgagcagaaggggaatggttacttgagaagaatggttccatttgg 1480
Db 241 tgaagcagaagaggtgagcagaaggggaatggttacttgagaagaatggttccatttgg 300

OY 1481 cattacatccactatagacagcctgtctccggtacacctgtgtgcatctgtggaagccac 1540
Db 301 cattacatccactatagacagcctgtctccggtacacctgtgtgcatctgtggaagccac 360

OY 1541 cagggtcgttatgaaagcactcgcagcagtgatatacaacggctgtccttcgcgaatttc 1600
Db 361 cagggtcgttatgaaagcactcgcagcagtgatatacaacggctgtccttcgcgaatttc 420

OY 1601 catgcgcagaagggcctctctggaagcagaagaatcttgcagaagtgagaanaaacacac 1660
Db 421 catgcgcagaagggcctctctggaagcagaagaatcttgcagaagtgagaanaaacacac 480

OY 1661 gcaagaanaattctgaagaagcataaacaatcattccttctgtagagtggtgcatcattata 1720
Db 481 gcaagaanaattctgaagaagcataaacaatcattccttctgtagagtggtgcatcattata 540

OY 1721 tcaagaagatcatccaaatacagccctgagccagaagaattggaagcttcttcaagtaa 1780
Db 541 tcaagaagatcatccaaatacagccctgagccagaagaattggaagcttcttcaagtaa 600

OY 1781 aagcttatatactaactaaggaagaatcccgattactatttggctcttcttgaacattt 1840
Db 601 aagcttatatactaactaaggaagaatcccgattactatttggctcttcttgaacattt 660

OY 1841 gcccaattgtgcaagtgctcgcgaactcatlaaactggaactttaaaggggagcattgac 1900
Db 661 gcccaattgtgcaagtgctcgcgaactcatlaaactggaactttaaaggggagcattgac 720

OY 1901 ttcatggaagaaagcctgcagagaagacaggtggaattccacatggaagagccccaagaa 1960
Db 721 ttcatggaagaaagcctgcagagaagacaggtggaattccacatggaagagccccaagaa 780

OY 1961 ctcaattcccgagggcgtgtatcttcttcttcaactggaagcagaagactcagaagact 2020
Db 781 ctcaattcccgagggcgtgtatcttcttcttcaactggaagcagaagactcagaagact 840

OY 2021 ggaagtcacactccggagtttcagcaagtgtgaataagcaaatatcacatatctcgggaa 2080
Db 841 ggaagtcacactccggagtttcagcaagtgtgaataagcaaatatcacatatctcgggaa 900
```

```
OY 2081 aatattcagctctgcacaaagcctcgaagctgcgaataaagaatgtgctggtgctgag 2140
Db 901 aatattcagctctgcacaaagcctcgaagctgcgaataaagaatgtgctggtgctgag 960

OY 2141 aagctcagttgtgtcctcgaagcctgtgaagaacttattctcctcagtggtggaagcag 2200
Db 961 aagctcagttgtgtcctcgaagcctgtgaagaacttattctcctcagtggtggaagcag 1020

OY 2201 tccctcacaataagaatggaagggcacatcacatctgttaacaaacctgaaacctgag 2260
Db 1021 tccctcacaataagaatggaagggcacatcacatctgttaacaaacctgaaacctgag 1080

OY 2261 tattcatgacctacagaatlcacaagcctgcgggtgtgtctgactgacagcttggttaact 2320
Db 1081 tattcatgacctacagaatlcacaagcctgcgggtgtgtctgactgacagcttggttaact 1140

OY 2321 gaagaaccttaacaaagctcacaatgtgataaacaataaagtgtgaatgagaagatgctataa 2380
Db 1141 gaagaaccttaacaaagctcacaatgtgataaacaataaagtgtgaatgagaagatgctataa 1200

OY 2381 actagctgaagcctggaanaacctgaagaagaatggttatttcaatttgaccactgtc 2440
Db 1201 actagctgaagcctggaanaacctgaagaagaatggttatttcaatttgaccactgtc 1260

OY 2441 tgaacttggagaggggaatggtattacatagtcgaagtcctgtcgaagtgaacctgtgacct 2500
Db 1261 tgaacttggagaggggaatggtattacatagtcgaagtcctgtcgaagtcgaacctgtgacct 1320

OY 2501 tgaagaanaattcaatagctcctgtctgtctgtctgtcgaatgcagtgaaatccctgacca 2560
Db 1321 tgaagaanaattcaatagctcctgtctgtctgtctgtcgaatgcagtgaaatccctgacca 1380

OY 2561 gaattctacaatttgcaccaacttgagcatcttgcattatcagaanaattacactggaanaa 2620
Db 1381 gaattctacaatttgcaccaacttgagcatcttgcattatcagaanaattacactggaanaa 1440

OY 2621 agatggaataatgaagcctctcattgaactatgcagaagatgaagtgctcagaacagctaac 2680
Db 1441 agatggaataatgaagcctctcattgaactatgcagaagatgaagtgctcagaacagctaac 1500

OY 2681 cgcactgtatgctgcgccttgaggctgtgagctgcaagcagcctggaacagcctgtgaaaca 2740
Db 1501 cgcactgtatgctgcgccttgaggctgtgagctgcaagcagcctggaacagcctgtgaaaca 1560

OY 2741 ttggagagaggtccccaacactcgtlcaagcttgggttgaanaacttgaagactcacagatac 2800
Db 1561 ttggagagaggtccccaacactcgtlcaagcttgggttgaanaacttgaagactcacagatac 1620

OY 2801 agagattgaattttaggtgcaatttcttggaagaagacctctgaaacctccagagctt 2860
Db 1621 agagattgaattttaggtgcaatttcttggaagaagacctctgaaacctccagagctt 1680

OY 2861 gaatttggcgggaatctgtgagcagtgatggaatggttgccttcagtggttgatttga 2920
Db 1681 gaatttggcgggaatctgtgagcagtgatggaatggttgccttcagtggttgatttga 1740

OY 2921 gaatttgaagcaatlaggttcttctgactttagtaactaaagaattctcactgataccagc 2980
Db 1741 gaatttgaagcaatlaggttcttctgactttagtaactaaagaattctcactgataccagc 1800

OY 2981 attagtcagaanaaccttagcacaagtgctatccaagttaacttcttgcaagaagctagact 3040
Db 1801 attagtcagaanaaccttagcacaagtgctatccaagttaacttcttgcaagaagctagact 1860

OY 3041 tgttggtggtcaatttgaatgtagatcgaatcgaatgtttaaaggttctttaaactagt 3100
Db 1861 tgttggtggtcaatttgaatgtagatcgaatcgaatgtttaaaggttctttaaactagt 1920

OY 3101 aactgttaataaagtgttactcgaagcagta 3133
Db 1921 aactgttaataaagtgttactcgaagcagta 1953
```


Db	2607	ATGCTGCGCCGTGGGGGCTGTGACGTGCAAGGACACCTCGAGACACTTGTGTAACAACTTTGGAG	2548
Qy	2148	gaggtccacaactcgttcaagcttgggtgtgaaaacctgagactccacagatcaagagatt	2807
Db	2547	GAGTCCCAACACCTCTCAAGCTTGGCTTGAAAACTCGAAGACTCAGACATACAGAGATT	2488
Qy	2808	agaattttagtgatcatttcttttggaaagaacccctcgtaaaactccagcagttgaattg	2867
Db	2487	AGAAATTTTAGGTGCATTTTGTGGAAAGAACCCCTCGAAAAAACTTCAGACAGCTGAATTTG	2428
Qy	2868	gcgsgaataatcgtltgagcagtgatgatatgcttcgctccatcagtggtgtatltgaaactc	2927
Db	2427	GCGGGAATTCGTGTGAGCAGTGTATGATGATGCGCTTGCCCTTCATGGGTGTATTTGAAATCTT	2368
Qy	2328	aagcaattagtglttcttttgaacttgaactgtactaagaagaatttcactgtatccagcatagtc	2987
Db	2367	AAGCAATTAAGTGTGTTTTGACTTTAGTACTTAAGAAATTTCTACCTGCATTCAGCATTTAGTC	2308
Qy	2388	agaaactatgcgaagtgtatccaaagtaactttctgcagaagctagcttgttggg	3047
Db	2307	AGAAATCTTGAAGCAAGTGTATTCCAAGTTAACTTTCTGCAAGAAGCTAGGCTTTGTTGGG	2248
Qy	3048	ttggaaattatgatatgatctcagtggttattacaggtgtctttaaactagaactgct	3107
Db	2247	TGGCAATTTATGATATGATATGATCTCAGTGTATTTACAGGTGCTTTTAACTAGTACTGCT	2188
Qy	3108	taataaagtgctactcgaagccagta	3133
Db	2187	TAAATTAAGTGTACTCGAAGCCAGTA	2162
RESULT 6			
AIID	AA114389	standard; DNA; 421 BP.	
AC	XX		
XX	AA114389;		
DT	XX		
DT	12-OCT-2001	(first entry)	
XX	XX		
DE	Probe #4322 for gene expression analysis in human cervical cell sample		
XX	XX		
KM	Probe: human; microarray; gene expression; cervical epithelial cell;		
KW	cervical cancer; ss.		
OS	Homo sapiens.		
XX	XX		
PN	W0200157278-A2.		
XX	XX		
PD	09-AUG-2001.		
XX	XX		
XX	30-JAN-2001; 2001WO-US00670.		
PF	XX		
PR	04-FEB-2000; 2000US-0180312.		
PR	26-MAY-2000; 2000US-0207456.		
PR	30-JUN-2000; 2000US-0608408.		
PR	03-AUG-2000; 2000US-0632366.		
PR	21-SEP-2000; 2000US-0234687.		
PR	27-SEP-2000; 2000US-0236359.		
PR	04-OCT-2000; 2000GB-0024263.		
XX	XX		
PA	(MOLE-) MOLECULAR DYNAMICS INC.		
XX	XX		
PI	Penn SG, Hanzel DK, Chen W, Rank DR;		
XX	XX		
DR	WPI; 2001-488901/53.		
XX	XX		
XX	Human genome-derived single exon nucleic acid probes useful for		
PT	analyzing gene expression in human cervical epithelial cells -		
XX	XX		
PS	Claim 25; SEQ ID No 4322; 487bp; English.		
XX	XX		
CC	The present invention relates to human single exon nucleic acid probes		
CC	(SNP). The present sequence is one such probe. The SNPs are derived		

CC from human HeLa cells. The SENPs can be used to produce a single exon
 CC microarray, which can be used for measuring human gene expression in a
 CC sample derived from human cervical epithelial cells. By measuring gene
 CC expression, the probes are therefore useful in grading and/or staging
 CC of diseases of the cervix, notably cervical cancer.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcr_sequences.

CC Sequence 421 BP; 126 A; 69 C; 77 G; 149 T; 0 other;

Query Match 7.7%; Score 242; DB 22; Length 421;
 Best Local Similarity 100.0%; Pred. No. 4.4e-61;
 Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2816 aggtgcatcttttggaaaaccctctgaaaactccagcaggtgaatttggcgggaaa 2875
 |||||||
 DB 180 aggtgcatcttttggaaaaccctctgaaaactccagcaggtgaatttggcgggaaa 239
 |||||||
 DB 2876 tcgttgagcagtgatgagtgcttgccttcatagggtgatttgaatacctaagaact 2935
 |||||||
 DB 240 tcgttgagcagtgatgagtgcttgccttcatagggtgatttgaatacctaagaact 299
 |||||||
 OY 2936 agtgttttttgacttagtaactaaagaatttctactgatacagcattagtcagaaaact 2995
 |||||||
 DB 300 agtgttttttgacttagtaactaaagaatttctactgatacagcattagtcagaaaact 359
 |||||||
 OY 2996 tagccaagtgatataccaagtaactttctgcaagaagctaggtgtgtggggaact 3055
 |||||||
 DB 360 tagccaagtgatataccaagtaactttctgcaagaagctaggtgtgtggggaact 419
 |||||||
 OY 3056 tg 3057
 ||
 DB 420 tg 421

RESULT 7

AAI35764
 ID AAI35764 standard; DNA; 421 BP.

AC AAI35764;

DT 17-OCT-2001 (first entry)

DE Probe #4450 used to measure gene expression in human placenta sample.

XX Probe: microarray; human; placenta; antenatal diagnosis;
 XX genetic disorder; ss.

OS Homo sapiens.

PN WO200157272-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00663.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0633366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 DR WPI: 2001-48897/53.
 XX Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human placenta -
 XX Claim 25; SEQ ID No 4450; 654bp; English.

CC The present invention relates to single exon nucleic acid probes (SENP).
 CC The present sequence is one such probe. The probes are useful for
 CC producing a microarray for predicting, measuring and displaying gene
 CC expression in samples derived from human placenta. The probes are useful
 CC for antenatal diagnosis of human genetic disorders.

CC Sequence 421 BP; 126 A; 69 C; 77 G; 149 T; 0 other;

Query Match 7.7%; Score 242; DB 22; Length 421;
 Best Local Similarity 100.0%; Pred. No. 4.4e-61;
 Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2816 aggtgcatcttttggaaaaccctctgaaaactccagcaggtgaatttggcgggaaa 2875
 |||||||
 DB 180 aggtgcatcttttggaaaaccctctgaaaactccagcaggtgaatttggcgggaaa 239
 |||||||
 OY 2876 tcgttgagcagtgatgagtgcttgccttcatagggtgatttgaatacctaagaact 2935
 |||||||
 DB 240 tcgttgagcagtgatgagtgcttgccttcatagggtgatttgaatacctaagaact 299
 |||||||
 OY 2936 agtgttttttgacttagtaactaaagaatttctactgatacagcattagtcagaaaact 2995
 |||||||
 DB 300 agtgttttttgacttagtaactaaagaatttctactgatacagcattagtcagaaaact 359
 |||||||
 OY 2996 tagccaagtgatataccaagtaactttctgcaagaagctaggtgtgtggggaact 3055
 |||||||
 DB 360 tagccaagtgatataccaagtaactttctgcaagaagctaggtgtgtggggaact 419
 |||||||
 OY 3056 tg 3057
 ||
 DB 420 tg 421

RESULT 8

AAI04213
 ID AAI04213 standard; DNA; 421 BP.

AC AAI04213;

DT 09-OCT-2001 (first entry)

DE Probe #4204 used to measure gene expression in human breast sample.

XX Probe: human; breast disease; breast cancer; development disorder; ss;
 XX inflammatory disease; proliferative breast disease; non-carcinoma tumour.

OS Homo sapiens.

PN WO200157270-A2.

PD 09-AUG-2001.

PF 29-JAN-2001; 2001WO-US00661.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0633366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 DR WPI: 2001-476286/51.
 XX

PT Novel single exon nucleic acid probe used to measuring gene expression
PT in a human breast -
XX
PS Claim 25; SEQ ID NO 4204; 322pp; English.
XX
CC The present invention relates to novel single exon nucleic acid probes.
CC The present sequence is one such probe. The probes are useful for
CC measuring human gene expression in a human breast sample, where the probe
CC hybridizes at high stringency to a nucleic acid expressed in the human
CC breast. The probes are useful for predicting, diagnosing, grading,
CC staging, monitoring and prognosticating diseases of the human breast,
CC particularly those diseases with polygenic aetiology. The diseases
CC include: breast cancer, disorders of development, inflammatory diseases
CC of the breast, fibrocystic changes, proliferative breast disease and
CC non-carcinoma tumours.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).
XX
SQ Sequence 421 BP; 126 A; 69 C; 77 G; 149 T; 0 other;

Query Match 7.7%; Score 242; DB 22; Length 421;
Best Local Similarity 100.0%; Pred. No. 4.4e-61;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2816 agtgcatttttgaagaacccctcgaacttcacagcagtgtaattggcgggaa 2875
DB 180 agtgcatttttgaagaacccctcgaacttcacagcagtgtaattggcgggaa 239
QY 2876 tcgtgtgacagtgatgatgcttcctcagtgatgattgaagaattcaagaact 2935
DB 240 tcgtgtgacagtgatgatgcttcctcagtgatgattgaagaattcaagaact 299
QY 2936 agtgtttttgacttagtactaagaattctcaccgtatccagcattagtcagaact 2995
DB 300 agtgtttttgacttagtactaagaattctcaccgtatccagcattagtcagaact 359
QY 2996 tagcacaagtgtatccaaagttaactttctcgaagaactgagctgtgtgtgtaact 3055
DB 360 tagcacaagtgtatccaaagttaactttctcgaagaactgagctgtgtgtgtaact 419
QY 3056 tg 3057
DB 420 tg 421

RESULT 9
AA123590
D AA123590 standard; DNA; 220 BP.
XX
AC AA123590;
XX
DT 12-OCT-2001 (first entry)
XX
DE Probe #13523 for gene expression analysis in human cervical cell sample.
XX
KW Probe: human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157278-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00670.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX
PS Claim 25; SEQ ID NO 13523; 487pp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENPs). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).
XX
SQ Sequence 220 BP; 58 A; 32 C; 54 G; 76 T; 0 other;

Query Match 7.0%; Score 220; DB 22; Length 220;
Best Local Similarity 100.0%; Pred. No. 9.1e-55;
Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2865 ttggcgggaactcgtgtgacagtgatgatgcttcctcagtgagtgatgaagaat 2924
DB 1 ttggcgggaactcgtgtgacagtgatgatgcttcctcagtgagtgatgaagaat 60
QY 2925 cttaagcattagtgctttttgacttagtactaagaattctcaccgtacccagcagcat 2984
DB 61 cttaagcattagtgctttttgacttagtactaagaattctcaccgtacccagcagcat 120
QY 2985 gtcaagaacttaagcaagtgatccaaagttaactttctgcaagaagcagtgctgt 3044
DB 121 gtcaagaacttaagcaagtgatccaaagttaactttctgcaagaagcagtgctgt 180
QY 3045 ggggtgcaattgatgatgatgatcctcggtttattacag 3084
DB 181 ggggtgcaattgatgatgatgatcctcggtttattacag 220

RESULT 10
AA148904
ID AA148904 standard; DNA; 220 BP.
XX
AC AA148904;
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #17590 used to measure gene expression in human placenta sample.
XX
KW Probe: microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 PI WPI; 2001-488897/53.
 DR
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human placenta -
 XX
 PS Claim 25; SEQ ID No 17590; 654bp; English.
 CC The present invention relates to single exon nucleic acid probes (SENP).
 CC The present sequence is one such probe. The probes are useful for
 CC producing a microarray for predicting, measuring and displaying gene
 CC expression in samples derived from human placenta. The probes are useful
 CC for antenatal diagnosis of human genetic disorders.
 XX
 SQ Sequence 220 BP; 58 A; 32 C; 54 G; 76 T; 0 other;
 Query Match 7.0%; Score 220; DB 22; Length 220;
 Best Local Similarity 100.0%; Pred. No. 9.1e-55;
 Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2865 ttggcggggaatcgtgtgagcagtgatgagtcgtccttcaatgggtatttggaaat 2924
 Db 1 ttggcggggaatcgtgtgagcagtgatgagtcgtccttcaatgggtatttggaaat 60
 QY 2925 cttaagcaattagtggttttttgaactttagtactaaagaatttctacctatccagcat 2984
 Db 61 cttaagcaattagtggttttttgaactttagtactaaagaatttctacctatccagcat 120
 QY 2985 gtcaagaacctagccaagtgtatcccaagttaactttctcgaagaagctaggttgt 3044
 Db 121 gtcaagaacctagccaagtgtatcccaagttaactttctcgaagaagctaggttgt 180
 QY 3045 ggggtgcaatttgatgatgatgatcctcagtggtattacaag 3084
 Db 181 ggggtgcaatttgatgatgatgatcctcagtggtattacaag 220
 RESULT 11
 AAI09206
 AAI09206 standard; DNA; 220 BP.
 AC AAI09206;
 XX
 DT 09-OCT-2001 (first entry)
 XX
 DE Probe #9197 used to measure gene expression in human breast sample.
 XX
 KW Probe; human; breast disease; breast cancer; development disorder; ss;
 KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
 XX
 OS Homo sapiens.
 OS
 PN WO200157270-A2.
 PD
 XX 09-AUG-2001.
 PD
 XX 29-JAN-2001; 2001WO-US00661.
 PF
 XX 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 PI WPI; 2001-476286/51.
 DR
 XX
 PT Novel single exon nucleic acid probe used to measuring gene expression
 PT in a human breast -
 XX
 PS Claim 25; SEQ ID No 9197; 322bp; English.
 CC The present invention relates to novel single exon nucleic acid probes.
 CC The present sequence is one such probe. The probes are useful for
 CC measuring human gene expression in a human breast sample, where the probe
 CC hybridises at high stringency to a nucleic acid expressed in the human
 CC breast. The probes are useful for predicting, diagnosing, grading,
 CC staging, monitoring and prognosing diseases of the human breast,
 CC particularly those diseases with polygenic aetiology. The diseases
 CC include: breast cancer, disorders of development, inflammatory diseases
 CC of the breast, fibrocystic changes, proliferative breast disease and
 CC non-carcinoma tumours.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 220 BP; 58 A; 32 C; 54 G; 76 T; 0 other;
 Query Match 7.0%; Score 220; DB 22; Length 220;
 Best Local Similarity 100.0%; Pred. No. 9.1e-55;
 Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2865 ttggcggggaatcgtgtgagcagtgatgagtcgtccttcaatgggtatttggaaat 2924
 Db 1 ttggcggggaatcgtgtgagcagtgatgagtcgtccttcaatgggtatttggaaat 60
 QY 2925 cttaagcaattagtggttttttgaactttagtactaaagaatttctacctatccagcat 2984
 Db 61 cttaagcaattagtggttttttgaactttagtactaaagaatttctacctatccagcat 120
 QY 2985 gtcaagaacctagccaagtgtatcccaagttaactttctcgaagaagctaggttgt 3044
 Db 121 gtcaagaacctagccaagtgtatcccaagttaactttctcgaagaagctaggttgt 180
 QY 3045 ggggtgcaatttgatgatgatgatcctcagtggtattacaag 3084
 Db 181 ggggtgcaatttgatgatgatgatcctcagtggtattacaag 220
 RESULT 12
 AAF58252/C
 AAF58252 standard; DNA; 936 BP.
 ID AAF58252;
 XX
 AC AAF58252;
 XX
 DT 24-APR-2001 (first entry)
 XX
 DE Oligonucleotide D1835.
 XX
 KW Electron-transfer group; ETM; mismatch; genotyping;
 KW gene expression; ss.
 XX
 OS Synthetic.
 OS
 PN WO200107665-A2.
 PD
 XX 01-FEB-2001.
 PD
 XX 26-JUL-2000; 2000WO-US20476.
 PF
 XX 26-JUL-1999; 99US-0145695.
 PR

[illegible][illegible]

[illegible]

XX	RESULT 1A
XX	AAF58257/C
ID	AAF58257 standard; DNA; 936 BP.
XX	
AC	AAF58257;
XX	
DT	24-APR-2001 (first entry)
XX	
DE	Oligonucleotide D1954.
XX	
OS	Electron-transfer group; ETW; mismatch; genotyping; gene expression; ss.
XX	Synthetic.
PN	WO200107665-A2.
PD	01-FEB-2001.
PE	26-JUL-2000; 2000MO-US20476.
PR	26-JUL-1999; 99US-0145695.
PR	17-MAR-2000; 2000US-0190259.
PA	(CLIN-) CLINICAL MICRO SENSORS INC.
PI	Umek RM;
DR	WPI; 2001-159728/16.
PT	Nucleic acids containing electron-transfer group, useful as labels in hybridisation assays, e.g. for genotyping, allowing repeat analyses on a single surface
XX	
XX	Example 6; Page 127; 159pp; English.

XX The present invention relates to a composition comprising two nucleic
XX acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
XX Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;
SO

Query Match	3.0%;	Score 94.8;	DB 22;	Length 936;
Best Local Similarity	0.8%;	Pred. No. 2.3e-17;		
Matches	6;	Conservative 461;	Mismatches 313;	Indels 0;
			Gaps	0;

[illegible]

RESULT	15
AAF58259/C	
ID	AAF58259 standard; DNA; 936 BP
XX	
AC	AAF58259;

XX 24-APR-2001 (first entry)
XX Oligonucleotide D2004.
DE Electron-transfer group; ETM; mismatch; genotyping;
XX gene expression; ss.
XX Synthetic.
OS
XX
XX MO200107665-A2.
PN
XX
XX 01-FEB-2001.
PD
XX 26-JUL-2000; 2000MO-US20476.
PF
XX 26-JUL-1999; 990S-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
XX (CLIN-) CLINICAL MICRO SENSORS INC.
PA
XX
XX Umek RM:
XX
XX WPI; 2001-159728/16.
DR
XX
XX Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
XX
XX Example 6; Page 128; 159pp; English.
PS
XX
XX The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
XX Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other:

Query Match 3.0%; Score 94.8; DB 22; Length 936;
Best Local Similarity 0.8%; Pred. No. 2.3e-17;
Matches 6; Conservative 461; Mismatches 313; Indels 0; Gaps 0;

DB 2341 taatgatacaataagaatgaagaagaatgataaaactagctgaagcctgaaa
Y 787
DB 2401 acctgaagaagatgtttattcatattgacccactgtctgacattgagaggaaatg
Y 727
DB 2461 atacatagatcaagtcctctgcaagtgaaacctgtgacctgaagaatcaattagct
DB 667
Y 2521 cctgctgctgtctgcaaatgcaatgcaaatcctagctcagaatctcacaatttgta
DB 607
Y 2581 aactgagactcttgatllatcagaataactcgtgaaagaatgaaatgaaatgcttc
DB 547
Y 2641 atgaactgacgaagaatgaacgtgctagaacagctcacgcactgactgctgctg
DB 487
Y 2701 gctgtgacgtgcaagcagcctgagcctgttgaaacattggaagagtgccacaac
DB 427

QY 2761 tcgtcaagcttggttgaaaactggagactcacagatacagagattagtg 2820
DB 367
QY 2821 catllttgaaagaacctctgaaaactcagcagctgaaattggcgggaatcgt 2880
DB 307
QY 2881 tgaagcagtgatgagctgctgctcctcagtggtgatttgagaatcctaagaattggt 2940
DB 247
QY 2941 ttltgacttagtaactaaagattctaccatgacagcattagtcagaaactagcc 3000
DB 187
QY 3001 aagtgatccaaagttaactttctgcaagaagctagctgtggtggaattgag 3060
DB 127
QY 3061 atgatgctcagtggtattacaggtgctttaactagtaactgcttaataaagtga 3120
DB 67

Search completed: March 25, 2002, 11:44:40
Job time: 3264 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2002, 10:50:15 ; Search time 2764.94 Seconds

(without alignments)
12176.222 Million cell updates/sec

Title: US-09-697-089-1

Sequence: 1 cgctcagcccggtggaag.....aagtgactcgaagccagta 3133

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 segs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estda:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: gb_est1:*
11: gb_est2:*
12: gb_hic:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	551	17.6	552	13	AQ309404 CITBI-EI-
C 2	477.4	15.2	480	10	AV719179 AV719179
C 3	407.8	13.0	602	13	AZ360053 1M0103H11
C 4	364	11.6	364	10	AI263294 qS57b01.x
C 5	347	11.1	371	10	AV656315 AV656315
C 6	308.8	9.9	476	11	BG210375 RST29913
C 7	282.8	9.0	470	13	AO624020 HS_5378.B
C 8	249	7.9	261	10	AM337918 he12h11.x
C 9	225.2	7.2	404	13	AO889169 HS_2161.B
C 10	170	5.4	553	13	AQ320928 RPTC11-93
C 11	165.8	5.3	840	11	BF207840 601862546
C 12	162	5.2	509	10	AW418826 ha21e11.x
C 13	157.2	5.0	219	13	AQ283886 RPTC11-78
C 14	153.8	4.9	251	11	BF903662 IL2-MT018
C 15	146	4.7	630	13	AO112439 CIT-HSP-2
C 16	124.8	4.0	412	11	H25984 y156g07.r1
C 17	106	3.4	728	13	AZ720059 RPTC-24-8
C 18	93.4	3.0	429	13	AZ484615 1M0311P08
C 19	66.2	2.1	429	13	AZ484615 1M0311P08
C 20	55	1.8	635	13	B58691 CIT-HSP-201
C 21	49.2	1.6	668	13	AZ762115 1M0556N15
C 22	46.8	1.5	625	13	AZ614134 1M0442N17
C 23	44.8	1.4	546	13	AZ362463 1M0107N03
C 24	42.2	1.3	1023	10	BE306291 601103781
C 25	40.8	1.3	835	10	BE643259 Cr-12.8_F0
C 26	40.6	1.3	361	10	AM504276 UT-HF-BNO
C 27	40.2	1.3	783	11	BG563851 602584655
C 28	40	1.3	424	13	AQ638259 927P1-17A
C 29	40	1.3	518	10	AA669206 ad93b07.s
C 30	40	1.3	833	13	AZ533736 ENTCH45TF
C 31	39.8	1.3	939	11	BE887277 601508592
C 32	39.6	1.3	467	10	AM988289 u905f04.Y
C 33	39.4	1.3	363	10	BE544063 601069826
C 34	39.4	1.3	523	11	BF259322 HYSMEF001
C 35	39.4	1.3	549	11	BG109381 602280435
C 36	39.4	1.3	573	11	BF257885 HYSMEF001
C 37	39.4	1.3	611	11	BF253881 HYSMEF000
C 38	39.4	1.3	754	11	BF258252 HYSMEF001
C 39	39.4	1.3	766	11	BF266655 HY_CRA001
C 40	39.4	1.3	854	11	BE881131 601482001
C 41	39.2	1.3	759	10	AI729528 BMLGH1135
C 42	39.2	1.3	806	13	AQ491970 drosoph11
C 43	39.2	1.3	1101	13	AL106910 drosoph11
C 44	39	1.2	1201	13	CNS007JU
C 45	38.6	1.2	574	10	AW658668 95275 MAR

ALIGNMENTS

RESULT 1
AQ309404/c 552 bp DNA 22-DEC-1998
LOCUS CITBI-EI.2528J13.rf CITBI-EI Homo sapiens genomic clone 2528J13,
DEFINITION DNA sequence.
ACCESSION AQ309404 GI:4041438
VERSION AQ309404
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 552)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Baas,S., Linher,K., Golden,K.,
Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
Venter,J.C.
TITLE Use of a random human BAC End Sequence Database for Sequence-Ready
Map Building
JOURNAL Unpublished (1998)
COMMENT Other-GSSs: CITBI-EI-2528J13.TR
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.
FEATURES
Location/Qualifiers
1..552
/organism="Homo sapiens"

REFERENCE 1 (bases 1 to 602)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weis,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert length: 10000 Std Error: 0.00
Plate: 0103 row: H column: 11
Seq primer: CGTGTAAACGACGCGCCAGT
Class: plasmid ends
High quality sequence stop: 602.

FEATURES
source
1. 602
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUCG1M0103H11"
/clone_1lb="Mouse 10kb plasmid UUCG1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: pMD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (g114732114[9b]AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance.

BASE COUNT 139 a 153 c 139 g 171 t
ORIGIN

Query Match 13.08; Score 407.8; DB 13; Length 602;
Best Local Similarity 80.3%; Pred. No. 2e-97;
Matches 478; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

OY 1171 aagctgtgctgcgaagtgactcctcggagcctgcgagcctgtggagactcgtcgg 1230
DB 595 AGGTGGAGGCTTCAGGTGATTTCACAGGAGCTGAGACTGTGGAGCTGCGCTAG 536
OY 1231 agggctgtctcccaagaattgattcgaactgcgagatgtgccagctgaatgag 1290
DB 535 AAGGTGTGTTCGCCACAAATTTGATTGTAACCCGAGCATGGGTCCAGCATGAACGAG 476
OY 1291 atgtctgtgacaaactgcctcctgttaataacagctcaaaagttcaagccaagt 1350
DB 475 ACGTCTGTGTGACAAATAGGCTCTCTGTAGTACACAGCTCAGAGGCTGAAGCCAGT 416
OY 1351 ataattcttcacaaagtcattcagagagatcacagagcgagagcagctgacgttat 1410
DB 415 ATAAATCTTTCATTAATCATTTTCAGAGATACAGCGAGGTGGAGACTCAGCAGTTTGC 356

OY 1411 tgacctcatalgaccagaagagtgaccacaagggaatggtacttgcagaaaaatggtt 1470
DB 355 TGAGCTCAAAAGAGCAGACAGAGGTGACCAAGGAAAGAGCTACTTAACAATGTTT 296
OY 1471 ccaattcgacattacatccattatagacccgtctcggtaaacctgtggatcgt 1530
DB 295 CCATCTCTGACATCATACATCCCTATATAGCATCTGCTCTACACGCTGGGTGTCGA 236
OY 1531 tggagaccaccagggctgttatgagacccctcgagcaggtatcacaagggctctc 1590
DB 235 CAGAAAGCAACCGGCGGTCTCATGAGGCCTTGCAGAAATGTTTATCAGACGCGACCTAC 176
OY 1591 tcgagcttcctatgcacagagggcctctctgagacaagaaatcttgaagttga 1650
DB 175 AAGGACTTTACGTACCAAGAGGCTCTCTGTGAGGCAAGATTCATCAGAGTCTGAGA 116
OY 1651 aacacctgagcaagaatcttgaagacataacatcaatccttcttgatgagttgca 1710
DB 115 ATACCAGTACGACAAAGATTTCTGAAGCCATCAATGTAATTCCTGCTGAGGTGCA 56
OY 1711 tccattatatacagagagatcaccatccaatcagccctgagccaagaattgaagc 1765
DB 55 TCAATTTGTTCTCAGAGAGTATGCTTAATCAGACCTGAGCCAGAAATTTGAAGC 1

RESULT 4
LOCUS AI263294/c 364 bp mRNA EST 03-FEB-1999
DEFINITION qx57b01.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2005417 3',
mRNA sequence.
ACCESSION AI263294
VERSION AI263294.1 GI:3871497
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 364)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Life Technologies catalog #: 11548-013
DNA Sequencing by: Washington University Genome Sequencing Center
clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/dbp/IMAGE/IMAGE.html
Insert length: 2146 Std Error: 0.00
Seq primer: -40UP from G1bco
High quality sequence stop: 364.

FEATURES
source
1. 364
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2005417"
/clone_1lb="NCI_CGAP_Pan1"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/note="Organ: pancreas; Vector: PCMV-SPORE6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.72 kb. Life Technologies catalog #:
11548-013"

BASE COUNT 117 a 84 c 55 g 108 t
ORIGIN

Query Match 11.68; Score 364; DB 10; Length 364;
Best Local Similarity 100.0%; Pred. No. 8.2e-86;
Matches 364; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2766 aagcttggttgaaactggaagactcacagatcacagagatttagtgatctt 2825

|||||
Db 364 AAGCTTGGCTGAAAAACGAGACTACAGATACAGAGATTAGATTAGTGCATT 305
OY 2826 ttgggaagaacccctcgaagaactccagcagttgaatttcgagggaatcgtggaac 2885
Db 304 TTTGGAAAGAACCCCTCTGAAAACTCCAGCAGTTGAAATTTGGCGGAATCGTGACC 245
OY 2886 agtgaatgaaggcttcgcttcacatgagtgatctgaagaacttaagaactagttctt 2945
Db 244 AGTGAATGATGGCTTGCCCTTCATGGGTATTTGAGAACTTAAGCAATTAGGTTT 185
OY 2946 gactctagctacaaagaattctctacatgacgaactatgtaagaactagccaagt 3005
Db 184 GACTTATGACTAAAGAAATTTCTACTGATCCAGCATTAAGTCAAGAACTTAGCCAAAGTG 125
OY 3006 ttatccaagttaactcttcctgcaagaagctagagctgttggttggaacttaagaat 3065
Db 124 TTAATCAATTAATCTTTTCGACAGAGCTAGGCTTGTTGGTGGCAATTTGATGATGAT 65
OY 3066 gactcagtgctattacaggtgctttaaactagtaactagctcttaataaagtgtactga 3125
Db 64 GATCCAGAGTGTATTTACAGTGCTGCTTTAACTAGTAACTGCTTAATAAAGTGCTACTCA 5
3126 agcc 3129
|||||
Db 4 AGCC 1
RESULT 5
LOCUS AV656315 371 bp mRNA EST 07-SEP-2000
DEFINITION AV656315 GLC Homo sapiens cDNA clone G1CE0A10 3', mRNA sequence.
ACCESSION AV656315
VERSION AV656315.1 GI:9877329
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 371)
Qian,B., Wu,T., Huang,Q., Huang,C., Kang,B., Gao,X., Xu,Z., Xiao,H.,
Zeng,L., Li,N., Peng,Y., Liu,F., Qu,J., Song,H., Cheng,Z., Qu,J.,
G., Yang,Y., Gu,Y., Tu,X., Jia,J., Fu,G., Ren,S., Zhong,M., Lu
Zeng,L., Xu,S., Gu,W., Chen,Z. and Han,Z.
TITLE Homo sapiens CDNA clone
JOURNAL Unpublished (2000)
COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919 (ex. 45)
Fax: 86-21-50801922
Email: hanzgchgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES
source
1..371
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="G1CE0A10"
/clone_lib="GLC"
/cissue_type="corresponding non cancerous liver tissue"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 112 a 85 c 91 g 82 t 1 others
ORIGIN

Query Match 11.1%; Score 347; DB 10; Length 371;
Best Local Similarity 98.6%; Pred. No. 2.8e-81;
Matches 350; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1908 gaaaagctgcgaagaacacaggtggaatccacatgaaagagccccaagaactacatt 1967
Db 1 GAAAAGCGCTGCAGAAACACAGAGTGAATCCACATGGAAGAGGCCCAAAACCTACATT 60
OY 1968 cccaagcagggctgtaactcttcttctcaactggaagcagaacttcgaacttgagtc 2027
Db 61 CCCAGCAGGGCTGTATCTTGTCTTCAACTGGAACAGAGAAATTCAGGACTCTGGAGGTC 120
OY 2028 acactcgggatttcagcaagtgaataagcaagatcacatccttcggggaataatc 2087
Db 121 ACACCTCCGGGATTTTCAGCAAGTTGAATTAAGCAGATATACATATCTGGGAAAAATATTC 180
OY 2088 agctctgcacaaagccttcagcgtcgcacaataaagatgctgctggtgctggaagcctc 2147
Db 181 AGCTTGCCACAAAGCCTCAGGCTGCMAATTAAGATGTGTGTGTGCTGGAAGCCTC 240
OY 2148 agttggtcctcagcagcctgtaagaacttatctctcaatgctggaagcagctccctc 2207
Db 241 AGTTGTGCTCTCAGCACCTGTAAAGCAATTTATCTCTATGCTGTGACACCCAGTCCCTC 300
OY 2208 accatagaagatgagagagcacatcacatctgtaacaacactgaaacacttgagta 2262
Db 301 ACCATAGAAAGATGAGAGGACACATCATCTGTACAAACCTGGACAGCTTGGGTA 355
|||||

RESULT 6
LOCUS BG210375 476 bp mRNA EST 21-APR-2001
DEFINITION RST29913 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION BG210375
VERSION BG210375.1 GI:13732062
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 476)
Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
Cain,S., Dahl,T., Thornton,M., Ramchandran,R., Whittington,J.,
Lerner,L., Krashoc,D., McElligott,K., Clark,S., Mays,R., Smith,E.,
Velloso,N., Hess,J., Cottrien,K., Lo,K., Offenbacher,J., Danzig,J.
and Ducart,M.
TITLE Creation of Genome-wide Protein Expression Libraries using Random
Activation of Gene Expression
JOURNAL Nat. Biotechnol. 19 (5), 440 (2001) In press
COMMENT Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scalc@atersys.com
High quality sequence stop: 360.

FEATURES
source
1..476
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="HT1080"
/note="See 'Creation of Genome-wide Protein Expression
Libraries using Random Activation of Gene Expression',
Nature Biotechnology, in press. Note that even though the
cell type indicated is HT1080, since a random activation
method was used, these sequence tags are not necessarily
expressed in HT1080 under normal circumstances."

BASE COUNT 120 a 99 c 107 g 149 t 1 others
ORIGIN

Query Match 9.9%; Score 308.8; DB 11; Length 476;
Best Local Similarity 97.8%; Pred. No. 4.2e-71;
Matches 313; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
OY 2811 atttagtgatcttttggaaagaccctcgaagaactccagagctgaatttgagc 2870

```
Db 145 ACTATAGTCATTTCTTGTGAAGAACCCCTGAAAAAATTCAGCAGTGAATTGCGC 204
QY 2871 ggaatcgtgtgacgagtgatgagtgctgctcctcatggtgttatttggaatcctaa 2930
Db 205 GGAATTCGTGTGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 264
QY 2931 caattagtgctttcttacttacttacttacttacttacttacttacttacttacttact 2990
Db 265 CATTAGTGTGTTTGTGCTTACTTAAAGAAATTTCTTACTTATCCAGCATTAAGTCAGA 324
QY 2991 aaacttagcgaagtgatcctcaagtgatccttctcaagtgatccttctggtggtg 3050
Db 325 AAACCTAGCCAAAGTGTATCCAGTTAACTTTCTGCAAGAACCTAGGCTTGTGGGTG 384
QY 3051 caattgagatgagatgagatcctcaagtgatccttctcaagtgatccttctggtggt 3110
Db 385 CATTAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 444
QY 3111 ataaagtctactcgaagcca 3130
Db 445 ATAAAGTGTACTCGAAGCCA 464

RESULT 7
LOCUS AO624020 470 bp DNA GSS 16-JUN-1999
DEFINITION HS-5378-B2.C12-SP6E RPCI-11 Human Male BAC Library Homo sapiens
ACCESSION AO624020
VERSION AO624020.1 GI:5086412
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 470)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieterdejong.med.bufileo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.bufileo.edu/ordering_bac.htm)
or from Research Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
plate: 954 row: F column: 24
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 470.
Location/Qualifiers
1..470
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=954 Col=24 Row=F"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACe3.6; Site.1: EcoRI; Site.2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACe3.6 vector at EcoRI sites"
```

```
BASE COUNT 142 a 101 c 107 g 117 t 3 others
ORIGIN
Query Match 9.0%; Score 282.8; DB 13; Length 470;
Best Local Similarity 84.1%; Pred. No. 3.5e-64;
Matches 317; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 2007 gaattcagagctctggagcgcacatccggagatttcgaagtgatgaagaagatc 2066
Db 2 GATATTCATGACTCTGGAGGTCCACACTCCGGGATTTTCAGCAGTNGAATTAAGCAAGATATC 61
QY 2067 acatattcggggaataatcattcagctctgcacaaagcctcagcgtcgtcaataaagaagct 2126
Db 62 AGATATCTGGGGAATAATTCAGCTCTGCCACAAAGCTCAGGCTGCCAATAAAGAGATGT 121
QY 2127 gctgtgtgtgctggaaagcctcagctgtgttcctcgaagcctgtlaagaacattatctc 2186
Db 122 GCTGTGTGTGCTGGAAAGCCCTCAGTTGTGCTCCTCAGCAGCTGTAAAGACATTTATTCCTC 181
QY 2187 atgtgtgaagccagtcctccacacatgaagaatgagagggcacatcattgttaacaac 2246
Db 182 ATGATGGAAGCCAGTCCCTCACCATTAGAAAGATGAGAGCAGATCATCTGTACACMAC 241
QY 2247 ctgaaaccttgatctacacacacacacacacacacacacacacacacacacacacacac 2306
Db 242 CTGAAACCTTGAGTATTCATGACCTTACAGATCAACAGCGCTCGCGGTATTTGTAATATC 301
QY 2307 agcttggttaacttgaagaaccttacaagctcataatgtgataacataaagatgaatgaa 2366
Db 302 AGTGTTGTGTGCTTTGTTCACTTAACATTAATAACAGTATTAACCTTGGAATAAGAGA 361
QY 2367 gaagatgtctataaac 2383
Db 362 GAGGAGACTTTATCTCT 378

RESULT 8
LOCUS AM337918/c 261 bp mRNA EST 31-JAN-2000
DEFINITION he12h11.x1 NCI-CGAP_CML1 Homo sapiens cDNA clone IMAGE:2918853 3',
ACCESSION AM337918
VERSION AM337918.1 GI:6834544
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 261)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaops-femail.nih.gov
Tissue Procurement: Elisabeth Paietta, Jonathan D. Licht, M.D.,
Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life
Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The
I.M.A.G.E. Consortium DNA Sequencing by: Washington University
Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bdrp/image/image.html
Seq primer: -400P from G1bco
High quality sequence stop: 201.
Location/Qualifiers
1..261
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2918853"
/clone_lib="NCI-CGAP_CML1"
/tissue_type="myeloid cells, 18 pooled CML cases, BCR/ABL"
```

rearrangement positive, includes both chronic phase and myeloid blast crisis"
/lab_host="DH10B"
/note="Organ: whole blood; Vector: pCMV-SPORT6; Site: 1; Salt: Site_2; Notif: Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies."
BASE COUNT 86 a 54 c 38 g 83 t
ORIGIN

Query Match 7.9%; Score 249; DB 10; Length 261;
Best Local Similarity 100.0%; Pred. No. 3,1e-55;

Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2885 cagatgagatgagctgctcctcatcagtgatcttgaagaatcctgaagaatgattttt 2944
|||
DB 261 CAGTATGATGAGCGCTGCTTCATCGGTATTTGAGATCTTAAGCATTTGTTT 202
|||
OY 2945 tgaattgactaagaattctacccatcagcattagtcagaaactagccaagt 3004
|||
DB 201 TGACTTTAGTACTAAGAATTTCTACCTGATCCAGCATTTAGCAAGT 142
|||
DB 3005 gttatccagtaactttctgcgaagaagctaggtctgttggtgcatttgatgata 3064
|||
DB 141 GTATCCAGTTACTTTCTGCAAGAGCTAGGCTTGTGGTGCAATTTGATGATCA 82
|||
OY 3065 tgaatcagctgcttatacagtgctcttaaacatgaactgcttaataaagctgactcg 3124
|||
DB 81 TGATCTAGTGTATTTACAGTGCTTTTAACCTAGTACTGTTAAATTAAGTACTGAC 22
|||
OY 3125 aagcagcta 3133
|||
DB 21 AAGCCAGTA 13

RESULT 9
AO889169/c 404 bp DNA GSS 10-NOV-1999
LOCUS HS_2161_B1_A01_77C CIT Approved Human Genomic Sperm Library D Homo
DEFINITION sapiens genomic clone Plate=2161 Col=1 Row=B, DNA sequence.
ACCESSION AO889169
VERSION AO889169.1 GI:6345359
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 404)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.

TITLE
JOURNAL
MEDLINE
COMMENT
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 2161 Row: B Column: 1
Seq primer: T7
Class: BAC ends
High quality sequence stop: 404.

FEATURES
SOURCE
Location/Qualifiers
1..404
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2161 Col=1 Row=B"

/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 119 a 93 c 70 g 122 t
ORIGIN

Query Match 7.2%; Score 225.2; DB 13; Length 404;
Best Local Similarity 91.2%; Pred. No. 7.2e-49;
Matches 250; Conservative 0; Mismatches 23; Indels 1; Gaps 1;

OY 34 caatgaattcataaaggacaataagccagagccctattcaagaatgggaatgactgtta 93
|||
DB 353 CAGTGAATTTCTATGAGACGACGATAGCCGAG -CCTTATTCGAAGATGGGAATGCTGT 295
|||
OY 94 taagaacaatccaatgacatgacatttgatgaaatgcttgatcgcgaagaagttaaca 153
|||
DB 294 TAGACTCTCTCAGATGACCTATTTGATGAATGATGATGAGAGAGAAATGAACA 235
|||
OY 154 tcaattgctgcgaagaagtgagagagatgctgctgctgagagagatcattacatgatttga 213
|||
DB 234 TCATTTTCTGAGAGAAAGTGGAGAGAGATGCTTTAGAGGATCTTACATGATTTTGA 175
|||
OY 214 aaaaaggctcagagtcctgttaacctcttcttaaaccttaagaagatggaactactc 273
|||
DB 174 AAGAGGGTTCAAGTCTGATACCTGTATGTTAAATCCCTTAAGAGAGAAATCTCTC 115
|||
OY 274 tattcagacttgatgacaaagcttttca 307
|||
DB 114 TATTTGAGACTTGAAGGACAAAGTAAGTATCA 81

RESULT 10
AO320928/c 553 bp DNA GSS 06-MAY-1999
LOCUS RPTC11-93C9.TV RPTC-11 Homo sapiens genomic clone RPTC-11-93C9, DNA
DEFINITION sequence.
ACCESSION AO320928
VERSION AO320928.1 GI:4053662
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 553)
Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
Use of human BAC End Sequences for Sequence-Ready Map Building
Unpublished (1998)
Other_GSSs: RPTC11-93C9.TV

TITLE
JOURNAL
MEDLINE
COMMENT
Contact: Shaying Zhao, William Niernan, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbeet@igr.org
Clones are derived from the human BAC library RPTC-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html
Seq primer: T7
Class: BAC ends.

FEATURES
SOURCE
Location/Qualifiers
1..553
/organism="Homo sapiens"
/db_xref="GDB:7535384"
/db_xref="taxon:9606"
/clone="RPTC-11-93C9"
/clone_lib="RPTC-11"

/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
RCC11 Human Male BAC Library"
BASE COUNT 170 a 107 c 114 g 162 t
ORIGIN

Query Match 5.4%; Score 170; DB 13; Length 553;
Best Local Similarity 100.0%; Pred. No. 3.7e-34;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2650 tcgacagatgaacgtgctctagaacagctcaccgactgctgcctggtgagc 2709
258 TCGACAGGATGAGCGCTAGAGACACTCACCGCACTGATGCTGCCCTGGGCGCTGAGC 139
2710 tcgaagcagcgtcagcagcgtctgaaacattggaagaggtcccaaacctgctgaac 2769
198 TGCACAGCGAGCTGACGACCTGTTGAAACATTTGGAGGAGGTCCCAACACTGCTCAAC 139
DB 198 TGCACAGCGAGCTGACGACCTGTTGAAACATTTGGAGGAGGTCCCAACACTGCTCAAC 139
QY 2770 ttgggtgaaactgagagctacagatagagattgaattagat 2819
138 TTGGGTGAAACTGAGACTCAGATACAGATTAAGATTTTAGGT 89

RESULT 11
BF207840 840 bp mRNA EST 06-NOV-2000
LOCUS BF207840/c 6018623461 NIH_MGC_53 Homo sapiens cDNA clone IMAGE:4082074 5',
DEFINITION mRNA sequence.
ACCESSION BF207840
VERSION BF207840.1 GI:11101426
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 840)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Clontech Laboratories, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LTCM944 row: 1 column: 11
High quality sequence stop: 636.

FEATURES
source
1..840
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4082074"
/clone_lib="NIH_MGC_53"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: bladder; Vector: pDNR-LIB (Clontech);
Site_1: SfiI (ggccgcctgcgc); Site_2: SfiI (ggccattgacgc
); Double-stranded cDNA was prepared from cell line RNA.
5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-ATCTAGAGGCGGAGCGCGGACATG-3' and 3' adaptor
sequence: 5'-ATCTAGAGGCGGAGCGCGGACATG-3' (30)BN-3'
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.55 kb (range 0.9-4.0 kb). 15/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)."

BASE COUNT
ORIGIN

276 a 143 c 165 g 256 t

Query Match 5.3%; Score 165.8; DB 11; Length 840;
Best Local Similarity 92.7%; Pred. No. 5.2e-33;
Matches 229; Conservative 0; Mismatches 12; Indels 6; Gaps 5;

QY 2888 tgatgatgctgctctcctcctcctggtgctatggaaccttaagaactagcttttga 2947
DB 679 TGATGATGCTCTCCCTCATGAGGTGA-TTGAANTCTTAACCATTAATGTT-TTGA 622
QY 2948 cttagtaactaagaattctaccctgacagcattagtaagaactagcc-aagtg 3006
DB 621 CTTTACTACTAAAGAAATTTCTACTCATCA-CATTAGTCTGAAACTTAGCCAAAGTGT 563
QY 3007 tatccagaatattcttcgaagaagcagcttgcttggtgctgaattgagatgag 3066
DB 562 TATCCAAAGTTAACTTTCTGCAAGAGCTAGGCT--CTTGGGTGCCATTTGATGATGATG 505
QY 3067 atctcagtgctatcagctgctttaaactagtaactgcttaactaagaagtgactcga 3126
DB 504 ATCTCAGTGTATTACAGGTGCTTTAACTACTACTGCTTAATTAAGTGTACTCGAA 445
QY 3127 gccagta 3133
DB 444 GCCAGTA 438

RESULT 12
AA418826/c 509 bp mRNA EST 09-FEB-2000
LOCUS AA418826 121111 NCI_CGAP_Kid12 Homo sapiens cDNA clone IMAGE:2874380 3',
DEFINITION mRNA sequence.
ACCESSION AA418826
VERSION AA418826.1 GI:6946758
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 509)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: Christopher Koskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 461.

FEATURES
source

1..509
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2874380"
/clone_lib="NCI_CGAP_Kid12"
/tissue_type="2 pooled tumors (clear cell type)"
/lab_host="DH10B"

/note="Organ: kidney; Vector: p773D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP_Kid12 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(clonoids 1323912-1325831, 1471368-1472903 and
1492104-1493255). Subtraction by Bento Soares and M.
Fatima Bonaldo."

BASE COUNT 161 a 99 c 80 g 169 t
ORIGIN

Query Match 5.2%; Score 162; DB 10; Length 509;
Best Local Similarity 100.0%; Pred. No. 4,9e-33;
Matches 162; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2972 tgcacagcattagtcagaaactagcgaagtgtatccaaagtaactttctgcaga 3031
|||||
Db 509 TGATCCACACTTAGTACGAAAACTAGCCAAAGTGTATCCAAAGTTACTTCTGCAGAA 450

3032 agctagcttctgtgtggtgcaatttgatgatgatctgctgtatccaaagcttt 3091
|||||
Db 449 AGCTAGGCTTGTGGGTGGCATTTGATGATGATGATCTCAGTGTATTACAGGTGCTTT 390

3092 taactagtaactgtcctaataagtgactcgaagccagta 3133
|||||
Db 389 TAACTAGTACTGCTTAATAAGTGTACTGCAAGCCAGTA 348

RESULT 13
LOCUS A0283886 219 bp DNA GSS 27-Apr-1999
DEFINITION RPII11-78E13, TV RPII-11 Homo sapiens genomic clone RPII-11-78E13,
DNA sequence.
ACCESSION A0283886
VERSION A0283886.1 GI:3910204
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 219)
Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
Berry,K., Granger,P., Suh,E., Wible,C., de Jong,P., and Venter,J.C.
Use of human BAC End Sequences for Sequence-Ready Map Building
Unpublished (1998)
COMMENT Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@igir.org
Clones are derived from the human BAC library RPII-11. For BAC
library availability, please contact Pieter de Jong
(pieter@jeng.med.buifalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buifalo.edu/ordering) or from
Research Genetics (info@resgen.com). BAC end search page:
http://www.cigr.org/tcd/humgen/bac_end_search/bac_end_search.html
Seq primer: T7
Class: BAC ends.

FEATURES
source
1. 219 Location/Qualifiers

/organism="Homo sapiens"
/db_xref="GDB:7529676"
/db_xref="taxon:9606"
/clone_id="RPII-11-78E13"
/clone_1id="RPII-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
RPII11 Human Male BAC Library"
BASE COUNT 69 a 45 c 46 g 59 t
ORIGIN

Query Match 5.0%; Score 157.2; DB 13; Length 219;
Best Local Similarity 86.1%; Pred. No. 8.1e-31;
Matches 174; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

2134 tggctggaagcctttagtctcagcaccctgtaagaacattattctctcatgtgtg 2193

Db 1 TGGCTGGAAGCCTTAGTGTGCTCCTACACACCTGTAGAACATTTATTCTCATGTGG 60
|||||

2194 aagccagtcacctcaccatagagaatgagagacatcacatctgtaacaaacctga 2253
|||||
Db 61 AAGCCAGTCCCTGACCATAGAGATGAGAGGACACATCATGTGTAAACAACTGAA 120

2254 ccttgatattcatgaactacagaaatcaagcgtctccggtgtgtctgactgacgttgg 2313
|||||
Db 121 CCTTGATATTTCATGACCTACAGATCAAGCGCTCCGGGTATGTGTATATCATGTGG 180

2314 gtaactgagaagacctaacaa 2335
|||||
Db 181 TGTGCTTTGTCATCTTAAAAA 202

RESULT 14
LOCUS BF903662 251 bp mRNA EST 18-Jan-2001
DEFINITION IL2-MT0180-181200-276-F03 MT0180 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF903662
VERSION BF903662.1 GI:12295121
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 251)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Negai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.R.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
2020263
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?IL2-IL2&t2=IL2-MT0180-
181200-276-F03&t3=2000-12-18&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 45
High quality sequence stop: 96.
Location/Qualifiers

FEATURES
source
1. 251

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="MT0180"
/dev_stage="Adult"
/note="Organ: marrow; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (O.S. Letters Patent application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
BASE COUNT 72 a 62 c 56 g 61 t
ORIGIN

Query Match 4.9%; Score 153.8; DB 11; Length 251;
Best Local Similarity 85.5%; Pred. No. 6.6e-30;
Matches 218; Conservative 0; Mismatches 32; Indels 5; Gaps 4;

218 tggctggaagcctttagtctcagcaccctgtaagaacattattctctcatgtgtg 2193

Oy	1042	agaccctcctctgtgtgcaatcaactgtgtgaattccaagtggct--gaagtgagttccac	1100
Db	1	AGACCCTCCTCTGTGAGTATCATTTTGTCAAATCTCTATGGACAAAAGGAATTCCAC	60
Oy	1101	tctcacacacaacaacgcgtgttccalacacctatagtactgttgtatacagaanaacaa	1160
Db	61	TCTCACACACAACAACGCAGTATCCATTACTTCTATGATCTGAATATACAGAAAAACAA	120
Oy	1161	cacaaacataaagttgtgtgcttcaagttagcttaattcaggagccttgagcaacgttgagac	1220
Db	121	CACAAACATAAAGCGCGTGTG--ATGTGACATCATTCGAGGCT-GACCACGTGTGG-GAC	176
Oy	1221	ctagctctggagggtgtgttcttcaccaaatgtgatlttgaactgcagagtgtgtcaccg	1280
Db	177	CTAAGTCTGGTGGGTGGCTTCCGCCAACAGTTGAATTCGAACCTGCAGAGTGTGCACCGC	236
Oy	1281	gtgaatgagatgttc	1295
Db	237	GCGTACTAGAATGTC	251

RESULT	15
LOCUS	AQ112439
DEFINITION	AQ112439 630 bp DNA
ACCESSION	CIT-HSP-2372C1.TR CIT-HSP Homo sapiens genomic clone 2372C1, DNA
VERSION	AQ112439
KEYWORDS	AQ112439.1 GI:3484599
SOURCE	GSS.
ORGANISM	human.
	Homo sapiens

REFERENCE AUTHORS

TITLE

JOURNAL
COMMENT

Unpublished (1998)
Other GSSs: CIR-HSP-22372C1.TF
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends.

FEATURES

source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2372C1"
/clone_1lib="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBELoBAC11;
HindIII"
BASE COUNT      188 a      126 c      124 g      192 t
ORIGIN

```

Query Match 4.7%; Score 146; DB 13; Length 630;
 Best Local Similarity 100.0%; Pred. No. 9.1e-28;
 Matches 146; Conservative 0; Mismatches 0; Indels 0;
 Gaps 0.

Accession	Sequence	Position
Db	485 AGCTGAAGCGCTGAAAAACCTGAGAGAGATGTGTTATTCATTGTGACCACCTGGTCTGA	544
Oy	244 cattggagaggaatgattacataagccaagtcctctctcaagagaacctgtaaccttga	250
Db	545 CATTGGAGAGGGAATGATTAACAATGCAAGTCTCTCTCAAGTGAACCCCTGTGACCTTGA	604
Oy	2504 agaaatcaatctagctctcctgctgct	2529
Db	605 AGAAATTCATTACTCTCTCTCTCTCT	630

Search completed: March 25, 2002, 11:37:02
Job time: 2807 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2002, 11:45:36 ; Search time 4399.74 Seconds

(without alignments)
11747.436 Million cell updates/sec

Title:

Sequence: 1 US-09-697-089-1
1 cgcctagccgcgtgggaag.....aagtgctcgaagccagta 3133

Scoring table:

OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1472140 seqs, 8248589755 residues

Word size: 0

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database:

Geneml: 1: gb_ba: 2: gb_htg: 3: gb_in: 4: gb_om: 5: gb_ov: 6: gb_pat: 7: gb_ph: 8: gb_pl: 9: gb_pr: 10: gb_ro: 11: gb_sts: 12: gb_sy: 13: gb_un: 14: gb_vl: 15: em_ba: 16: em_fun: 17: em_hum: 18: em_in: 19: em_om: 20: em_or: 21: em_ov: 22: em_pat: 23: em_ph: 24: em_pl: 25: em_ro: 26: em_sts: 27: em_sy: 28: em_un: 29: em_vl: 30: em_htgo_hum: 31: em_htgo_inv: 32: em_htgo_rod: 33: em_htg_hum: 34: em_htg_inv: 35: em_htg_rod: 36: em_htg_other:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	3133	100.0	3133	9	AY032589 Homo sapi
2	2946	94.0	3219	9	AY035391 Homo sapi
3	2915	93.0	3396	9	AY027787 Homo sapi
4	2912	92.9	3581	9	AF376061 Homo sapi
5	1946	62.1	160583	2	AC010968 Homo sapi
6	1844	58.9	138909	9	CNS01DS3
7	1290	41.2	1335	9	IR2003417
8	836	26.7	1395	9	AY027788 Homo sapi
9	485	15.5	768	9	AY027789 Homo sapi
10	318	10.2	162692	9	CNS01DS8
11	318	10.2	185281	2	AC011232
12	293	9.4	578	9	AY027790
13	170	5.4	535	11	G55568
14	25	0.8	160583	2	AC010968
15	23	0.7	175152	2	AC074195
16	23	0.7	183556	2	AC019059
17	23	0.7	188459	2	AC090582
18	23	0.7	204487	2	AC090559
19	22	0.7	87834	9	AP002957
20	22	0.7	151088	9	AC020917
21	22	0.7	151366	33	AC007951
22	22	0.7	153733	2	AC011980
23	22	0.7	155531	2	AL592213
24	22	0.7	156230	2	AP002958
25	22	0.7	159946	2	AC079194
26	22	0.7	186413	2	AL161913
27	22	0.7	194533	2	AL592438
28	21	0.7	7595	4	BTCASK35
29	21	0.7	37906	9	AC005199
30	21	0.7	45496	2	AC017948
31	21	0.7	76526	2	AC027055
32	21	0.7	80346	8	ATM4E13
33	21	0.7	84499	8	ATF23E12
34	21	0.7	86710	2	AL513468
35	21	0.7	130117	2	AC004907
36	21	0.7	157308	2	AC079351
37	21	0.7	183839	2	AC024590
38	21	0.7	184864	2	AC013553
39	21	0.7	194874	2	AC080090
40	21	0.7	195165	8	ATCHRIV82
41	21	0.7	195217	2	AC025566
42	21	0.7	197859	8	ATCHRIV83
43	21	0.7	198935	2	AC068573
44	21	0.7	206137	2	AL593857
45	21	0.7	211030	2	AL513468
46	21	0.7	221341	2	AC092992
47	21	0.7	227194	2	AC020727
48	21	0.7	303367	3	AE003538
49	20	0.6	768	9	AY027789
50	20	0.6	865	2	AC078439
51	20	0.6	1092	2	AC047412
52	20	0.6	1198	14	REOS3NSB
53	20	0.6	1688	8	SCYBR141C
54	20	0.6	2833	8	SCYBR142M
55	20	0.6	3709	3	AF116341
56	20	0.6	12595	8	SCIRAI
57	20	0.6	36631	9	HUMCOL7A1X
58	20	0.6	40937	3	CER01E6
59	20	0.6	45027	3	CHRG39N14
60	20	0.6	45459	2	AC006103
61	20	0.6	83969	9	AC005210
62	20	0.6	86155	9	AL159169
63	20	0.6	86719	9	AP000885
64	20	0.6	88326	3	AC005923
65	20	0.6	88839	3	AC084447
66	20	0.6	91733	2	AL391278
67	20	0.6	93409	2	AC073294
68	20	0.6	102185	2	AC083819
69	20	0.6	106601	9	AL357352
70	20	0.6	109810	9	AL139036

C 71	20	0.6	114169	9	AC010902	Homo sapi
C 72	20	0.6	120733	9	AC022124	Homo sapi
C 73	20	0.6	121931	9	AC008390	Homo sapi
C 74	20	0.6	124347	9	AC010072	Homo sapi
C 75	20	0.6	129837	9	AC004829	Homo sapi
C 76	20	0.6	130981	2	AC087702	Tyrpanoso
C 77	20	0.6	132171	3	AC008370	Drosophila
C 78	20	0.6	137955	2	AL359752	Human DNA
C 79	20	0.6	141079	8	AP002868	Oryza sat
C 80	20	0.6	142203	8	AC079521	Mus muscu
C 81	20	0.6	145013	2	AC010435	Homo sapi
C 82	20	0.6	145576	8	AP002541	Cryza sat
C 83	20	0.6	146545	2	AC090564	Homo sapi
C 84	20	0.6	147728	2	AC034113	Homo sapi
C 85	20	0.6	157559	2	AC020344	Drosophila
C 86	20	0.6	158574	2	AC024895	Homo sapi
C 87	20	0.6	159158	9	AL359915	Human DNA
C 88	20	0.6	161590	2	AC092727	Bos tauru
C 89	20	0.6	164785	2	AC027540	Homo sapi
C 90	20	0.6	166953	3	AC093104	Drosophila
C 91	20	0.6	168666	2	AC074324	Homo sapi
C 92	20	0.6	170279	2	AC027053	Homo sapi
C 93	20	0.6	173933	9	AP002022	Homo sapi
C 94	20	0.6	174098	9	AC005737	Homo sapi
C 95	20	0.6	175086	2	AL157889	Homo sapi
C 96	20	0.6	175213	2	CNS07EF6	Homo sapi
C 97	20	0.6	176039	2	AC073065	Homo sapi
C 98	20	0.6	177290	2	AC013536	Homo sapi
C 99	20	0.6	177916	2	AL589863	Homo sapi
C 100	20	0.6	178861	2	AC068231	Homo sapi

ALIGNMENTS

RESULT	1					
LOCUS	AY032589	3133 bp	mRNA	PRI	25-MAY-2001	
DEFINITION	Homo sapiens caspase recruitment domain protein 12 mRNA, complete cds.					
ACCESSION	AY032589	GI:13899172				
VERSION	AY032589.1	GI:13899172				
KEYWORDS	human.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 3133)					
AUTHORS	Geddes, B. J., Wang, L., Huang, W.-J., Lavellee, M., Manji, G. A., Brown, M., Junman, M., Morganstern, J., Merriam, S., Glucksmann, A., Distefano, P. S. and Berlin, J.					
TITLE	Human CARD12 is a Novel CED4/Apaf-1 Family Member That Induces Apoptosis					
JOURNAL	Biochem. Biophys. Res. Commun. 284 (1), 77-82 (2001)					
PUBMED	11374873					
REFERENCE	2 (bases 1 to 3133)					
AUTHORS	Berlin, J.					
TITLE	Direct Submission					
JOURNAL	Submitted (15-APR-2001) Neurobiology, Millennium Pharmaceuticals Inc., 640 Memorial Drive, Cambridge, MA 02139, USA					
FEATURES	Location/Qualifiers					
source	1..3133					
CDS	/organism="Homo sapiens"					
	/db_xref="taxon:9606"					
	36..3110					
	/note="CARD12: CED4/Apaf-1 family member"					
	/codon_start=1					
	/product="caspase recruitment domain protein 12"					
	/protein_id="AAK38730.1"					
	/db_xref="GI:13899173"					
	/translation="MNFIDNSRALIORMGWTIVKQITDLEFVNNVLRREVNIIICE KVEDDAAGIITHMLIKKSGSCNLEKSLKMNWPLDIDNGSLFPHOTSGDDIDA ODKLDYHTPSFLNFPYLGEDIDITFMKSTFTEPVLMRKDDHHRVQDLTLNGLDA					

Query Match	100.0%	Score 3133;	DB 9;	Length 3133;	
Best Local Similarity	100.0%	Pred. No. 0;			
Matches 3133;	Conservative	0;	Mismatches	0;	Gaps 0;
1	CGCTTACCCGCGTGGAGAGCTTCATCCAGAACATATTTCTAAAGGCAATFACC	60			
1	CGCTTACCCGCGTGGAGAGCTTCATCCAGAACATATTTCTAAAGGCAATFACC	60			
61	GAGCCCTATTCAGAAATGGAGATGACTGTTATTAACCAATCAATGACCTATTG	120			
61	GAGCCCTATTCAGAAATGGAGATGACTGTTATTAACCAATCAATGACCTATTG	120			
121	TATGGAATGTTCTGAATGCGAAGATTAACATCATTTGCGCAGAGGTGAGCAG	180			
121	TATGGAATGTTCTGAATGCGAAGATTAACATCATTTGCGCAGAGGTGAGCAG	180			
181	ATGCTGCTAGAGGATTCATCATGATTTTGAAGAGGTTCAAGTCTGTAACCTCT	240			
181	ATGCTGCTAGAGGATTCATCATGATTTTGAAGAGGTTCAAGTCTGTAACCTCT	240			
241	TCTTAAATCCCTTAAGAGTGAAGTCTCTCTATTCAGAGTCTGAAGTGAAGTCT	300			
241	TCTTAAATCCCTTAAGAGTGAAGTCTCTCTATTCAGAGTCTGAAGTGAAGTCT	300			
301	TTTTCATCAGACATCAAGAGGAGCTTGAGATTTGGCTCAGAGATTTAAAGGCTT	360			
301	TTTTCATCAGACATCAAGAGGAGCTTGAGATTTGGCTCAGAGATTTAAAGGCTT	360			
361	ACCATACCCATCTTCTGAACTTATCCCTTGGTGAAGATATGACATTTATTTTA	420			
361	ACCATACCCATCTTCTGAACTTATCCCTTGGTGAAGATATGACATTTATTTTA	420			
421	ACTTAAAGAGACCTTCAAGACCTGCTGAGAGGAGGAGGAGGAGGAGGAGGAG	480			
421	ACTTAAAGAGACCTTCAAGACCTGCTGAGAGGAGGAGGAGGAGGAGGAGGAG	480			
481	TGGAGAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	540			
481	TGGAGAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	540			
541	GAGAGAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	600			
541	GAGAGAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	600			
601	GAGAGAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	660			
601	GAGAGAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	660			

QY 661 aggttgacttttgaaccctctgtatcaactccctgtatatactctgtgcacatcaaga 720
|||||
Db 661 AGGTTGACTTTTGAACCCCTCTGTATCAACTCTGGATATACCTGGACATATCAGA 720
QY 721 agcagacatctatgscatctgtctgaagctgtgcagcagaaggtctcttctctgaty 780
|||||
Db 721 AGCAGACATTCATGAGCCTGCTGCTGAAGCTGCGCAGAGAGGTTCTTCTTCTTGATG 780
QY 781 gctacacatgaattcaagcccccaactgtcccaaaatcgagccctgtataaagaacc 840
|||||
Db 781 GCTACATATGATTCAGCCCAAGACCTGCCAATAATCGAAGCCCTGATTAAGGAACC 840
QY 841 accgtctcaagaacatctgtatctgtacacacacacacacacacacacacacacacac 900
|||||
Db 841 ACCGCTTCAAGAACATGCTCATCTGTCACCACTGATGCTCTGAGGCAATACGGC 900
QY 901 agtttgctccctgtactgtctgaagctgtgaggtatatagaagaagaacgcacacacac 960
|||||
Db 901 AGTTTGCTCCCTGACTGCTGAGCTGGGGATATGACAGAAACAGGCGCCAGCTCTCA 960
QY 961 tccgaagaatgtctgtatcaagaagacgtctgtcgaagctgtgtctcccaaatccaaatcca 1020
|||||
Db 961 TCCGAAGAAGTGTGATCAAGAGACTTGTGAAGGCTTGTGCTCAATTCAGAAATCCA 1020
QY 1021 gctgtctgaagaaatccatgaagaacccctctcttctgtgtcatcaactgtgtcatccaga 1080
|||||
Db 1021 GCTGCTTGAAGAAATCTCATGAACACCCCTCTCTTGTGTGCTCATCTGTGCAATCCAGA 1080
QY 1081 tgggtgaagaatgtctcaactctcacacacacacacacacacacacacacacacacacac 1140
|||||
Db 1081 TGGGTGAAGAATGTCTCAACTCTCACACACAAACAGGCTGTCCATCTCTATGATC 1140
QY 1141 tgtgtgtacagaanaaacaacaacaataaagtggtgtgtcaagtgatctatccgga 1200
|||||
Db 1141 TGTGTGTACGAANAACAAACAAACATTAAGGTGTGCTGCAAGTCTTATTCGGA 1200
QY 1201 gctctgaacacatctgtgagacatctgtcgtgaggtgtgtctcccaaatcttgatctg 1260
|||||
Db 1201 GCTGTAACCATCTGTGAGACACTGCTGTGAGGCTGTCTCTCCCAATTTGATTTCCG 1260
QY 1261 aactgcagaaatgtctccagcgtgaatgaagaatgtcctgtcgaacacacacacacacacac 1320
|||||
Db 1261 AACTGCAGGAATGTCTCAGAGGTGAATGAGGATGTCCGTCGACACACGCGCTCTCTGTA 1320
QY 1321 aatatcaactcaagaagttcaagaacaaatataatctcttcaagaatcttcaagaagt 1380
|||||
Db 1321 AATATACACTCAAGGTTCAAGCCAAAGTATTAATTTCTTCAAACTCATTTCCAGAGT 1380
QY 1381 acacagcagaaacgaagacatcaagcagttatctgaagctctcaagaagaagaagtgcga 1440
|||||
Db 1381 ACACAGCAGAACGACACTGACAGTTTATGTACGCTCATGAGCCAGAGAGGTGACCA 1440
QY 1441 aggggaatgttactctgcagaaataatgttctcaatctcgaacatctacacactataga 1500
|||||
Db 1441 AGGGGAATGTCTCTGCAAGAAATGTTCATTTCTGACATTTACATCCACTTATAGCA 1500
QY 1501 gctgtgtctgtacacactgtgtgtctgtgtgaagccacacagagctgtgtataaagcacc 1560
|||||
Db 1501 GCTGTGCTCCGTAACCTGTGGGTCACTGTGTGAAGCCACAGGCGCTGTATGAAGACC 1560
QY 1561 tctgcacagtgatctacacacagctgtcctctcgtgaactctcagcccaagaagcctctc 1620
|||||
Db 1561 TCTGCACAGTGTATCAACACGCTGCTTGTGAGCTTTCATCGCCAAAGGCGCTCTCT 1620
QY 1621 gtagacagaaatctctgtcaaatgtgtgaataaacacacacacacacacacacacacacac 1680
|||||
Db 1621 GGAGACAGGAATCTTGCAAAGTGTGAATAAACCACTGAGCAAGAAATCTGTAAGGCCA 1680
QY 1681 taaacatcaatctctctgtgtgaagtggtgtgtcatctatatacaagaaggaatccaat 1740
|||||
Db 1681 TAAACATCAATTTCTTGTATGAGTGTGCTCATTTATATCAAGAGATGATTCATAAT 1740
QY 1741 cagccctgagcagaatcttgaagcttcttcaaggttaaagcttatatacaactcag 1800
|||||

Db 1741 CAGCCCTGAGCCCAAGATTTGAAGCTTTCTTGAAGTAAACCTTATATATCACTCAG 1800
QY 1801 ggaacacacccagatctactatctgtactctctgaacatctgtcccaatctgtcagtgctc 1860
|||||
Db 1801 GGAACATCCCGAATTTCTTGTATGACTTCTTGAACATTTGCCCAATTTGTGAAGTGTCT 1860
QY 1861 tgaactcatctaaactgtgacttttatgtggagctataggtctcaatgtggaagaagctgtcag 1920
|||||
Db 1861 TGACTTCATTTAAACTGTGACTTTTATGGGGAGCTATGGCTTATGGGAAAGGCTTCAG 1920
QY 1921 aagacacagtggaatccatctgaagaagccccagaacacacacacacacacacacacacacac 1980
|||||
Db 1921 AAGACACAGTGTGAATCTCATGTGAAGAGGCCCCAGAAACCTTACATTTCCAGCAGGCTG 1980
QY 1981 tatcttctctcaactgtgaagaagaatctgaagcctgtgaagctgtacacacccggat 2040
|||||
Db 1981 TATCTTGTCTTCAACTGTGAAGAGGAATTCAGAGACTGTGAGGCTCACACTCGGGATT 2040
QY 2041 tcaagaagttggaatgaagaatatacatatctgtgggaataatataagctctgtccaca 2100
|||||
Db 2041 TCAGCAAGTTGAATTAACCAAGATATCATATCTGGGAAATAATTCAGCTGTGCCCAA 2100
QY 2101 gctcaagctcgaataaagaatgtgtgtgtgtgtgaagcctcagctgtgtccca 2160
|||||
Db 2101 GCCTCAAGCTGCAATTAAGAGATGTCTGTGTGTGTGTGAAGCTTCAAGTTGGTCTCA 2160
QY 2161 gcaactgtgaagaatctatctctcatgtgtgaagcagctccctccacataagaatgt 2220
|||||
Db 2161 GCACCTGTGAAGCAATTTATCTCTCATGTGTGAAGCAGCTCCCTCCCATGAAAGATG 2220
QY 2221 aagagcacaatcatctgtatacaaacacacacacacacacacacacacacacacacacacac 2280
|||||
Db 2221 AGAGGCATCATCTCTGTAAACAACTGAAACCTTGAAATTCATGACCTTACAGATC 2280
QY 2281 aacgctgtccgggtgtgtctgtacgtgacagctgtgtgaactgtgaagaacacacacacacacac 2340
|||||
Db 2281 AACGCTGCCGGGTGTGTCTCATGTGACAGCTGTGGTAACTTGAAGAACCTTACAAAGCTCA 2340
QY 2341 taatgataacataaagaatgaatgaagaatgtctataaactagtgtgaagcctgaataa 2400
|||||
Db 2341 TAATGATTAACATTAAGATTAAGAAAGATGATCTATTAACATGAGAGGCTGTGAAA 2400
QY 2401 aactgaagaagaatgtgtttatctatctgaacacacacacacacacacacacacacacacacacac 2460
|||||
Db 2401 AACTGAAGAAGATGTGTATTTATTTGATTTGACCACTGTGTGACATTTGAGAGGAAATGG 2460
QY 2461 attacatagtcgaagctctgtcaagtgaaacacacacacacacacacacacacacacacacacac 2520
|||||
Db 2461 ATTACATAGTCAGATGTCTCTCAAGTGAACCTGTGACCTTGAAGAAATTCATTTACTCT 2520
QY 2521 cctgtgtctgtctgtcaaatgtcagtgaaatcctagctcagaatctcacaatctgtca 2580
|||||
Db 2521 CCTGTGCTGTGTGTGCAAAAGCAGTGAATAATCTAGCTCGAATCTTCACAATTTGGTCA 2580
QY 2581 aactgaacacatctgtgtttatctagaataaactctgtgaagaagaatgtgaagaagctctc 2640
|||||
Db 2581 AACTGACATTTCTTGATTTATTCAGAAATTAACCTGTGAAAGATGGAAGAAAGCTCTTC 2640
QY 2641 atgaacatctgaacagatgaacgtgtcagaacacacacacacacacacacacacacacacacacac 2700
|||||
Db 2641 ATGAACATGATCGAAGAGATGAACGTGTGTAACACAGCTCACCGACATGATCTCGCTCGG 2700
QY 2701 gctgtgaagtgcaagcagcctgtgagcctgtgtgaacatctgtgaggaaggtccacacac 2760
|||||
Db 2701 GCTGTGACGTGCAAGGAGCCTGAGCAGCTGTGTGAATAATTTGTGAGAGAGTCCCAAC 2760
QY 2761 tctgaagctgtgtgtgaagaactgtgagactcagaataagaagatgaagattgaagtg 2820
|||||
Db 2761 TCGTCAAGCTTGGTGTGAATAACTGTGAGACTCACAGATACAGAGATTAATTAATTTAGGTG 2820
QY 2821 catcttctgtgaagaacccctcgaagaactcagcagctgtgaattgtgcgggaatcgt 2880
|||||

Db 2821 CATTTTTGGAGAAAGAACCTCTGTAACCAAGCAGTTGAATTTGGCGGGAATTCGTG 2880
 QY 2881 tgaagcagatgatgagcctgctgctcctcaatgagtgatattgagaactcttaagcaattagtg 2940
 Db 2881 tgcagcagatgatgagcctgctgctcctcaatgagtgatattgagaactcttaagcaattagtg 2940
 QY 2941 ttttgacttgactaagaattctcactgacatccagcattagtcagaaacttagcc 3000
 Db 2941 TTTTGACTTTAGTACTAAGAAATTTCTACCTGATCCACATTAAGCAGAAACTTAGCC 3000
 QY 3001 aagtgatccagatgaactcttctgcaagaagcagcctgctggtgagcaattgagtg 3060
 Db 3001 AAGTGTATTCACAAAGTTAACTTTTTCGCAAGAGCTAGGCTTGTGGTGCAATTTGATG 3060
 QY 3061 atgacgacatcagtgcttattacaggtgcttcaactagtaactcttaataaagtga 3120
 Db 3061 ATGATGATCTCAGTGTATTATACAGGTGCTTTTAACTAGTAAGCTTAATAATTAAGTGA 3120
 QY 3121 ctgcaagccagta 3133
 Db 3121 CTGCAAGCCAGTA 3133
 RESULT 2
 LOCUS AY035391 3219 bp mRNA PRI 24-JUL-2001
 DEFINITION Homo sapiens ICE-protease activating factor mRNA, complete cds.
 ACCESSION AY035391
 VERSION AY035391.1 GI:14334214
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 3219)
 Poyet,J.L., Srinivasula,S.M., Thnani,M., Razmarz,M.,
 Fernandes-Alnemri,T. and Alnemri,E.S.
 Identification of Ipaf, a human caspase-1-activating protein
 related to Apaf-1
 J. Biol. Chem. 276 (30), 28309-28313 (2001)
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 21359454
 11390368
 2 (bases 1 to 3219)
 Poyet,J.-L., Srinivasula,S.M., Fernandes-Alnemri,T. and
 Alnemri,E.S.
 Direct Submision
 Submitted (16-MAY-2001) Microbiology and Immunology, Thomas
 Jefferson University, 233 S. 10th Street, Philadelphia, PA 19107,
 USA
 FEATURES
 Source
 Location/Qualifiers
 1. 3219
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /map="821-22"
 145. 3219
 /note="IPAF: CED4/Apaf-1 family member; caspase-associated
 recruitment domain containing protein"
 /codon_start=1
 /product="ICE-protease activating factor"
 /protein_id="AAK59843.1"
 /db_xref="GI:14334215"
 /translation="MNFTRKNSRLIORMGMTVKTITDDLFVNVNLRBENIICE
 KVEDDAAGIILHMLLKSGESCNLFSLKSMNPLFDLNGOSLPHQTSGLDDILA
 ODLDLHTPEFLNEVPLGEDIDILFNKSTPEVPLMRKDNHRRVQLNLNGILOA
 LSPCITGESGSGKSTLRIAMIMSGSKALFKFVPEFLISRAOSGLPFTICD
 OLDPITKIKOTPMAMLIKLRQYFLLDGYNEFRPQNCPEIEALIKENIRKMWI
 VTTTECLRHIFOGALTAEGDMEDESAQLIREVLLKELAEGLLQIQSRCLRN
 MKPLFVITCAIOMGESEFHSHTOTLFHTFYDLLLQKNHKKHGAASPLISID
 CGDLAEVFSHKEDFELQDVSYNEDVLTGLLCTTAQRKRYKFFKRSDEYT
 AGRRSLSLTSHPEEVTKNGYLOKNVYSIDITSTYSKLTLYTCGSSVEATRAVMK
 LAAYVHGCLIGLSTAKRPLRQESLOSVKNTTDEILKALININSEVFCGHIVQEST
 SKSLSDQEFAPQOKSLYINSGNIPDLTFEHLPCASALDELKIDFYGAAAMW

EKAEDTGGIIMEAPETYIPBSRAVSLFFNNKQERFLEVLTRDFSKLNKODIRYLG
 IESSATSLRQICRQAGVAGSLIVLSTCKNYSIMVASEPILTEDEKHTSVNFKT
 LSIHLOMORLPGLGTDLSGNKNTLXIMINIKNEDAKIKLEKNTKMKCLFHL
 THLSIDIGDMOYIYKSLSSPCDDEIDIVSCCLSANNAVKTLAONLNUYKSLIDS
 ENYLEKDENEALHIDIRMNVLDEDTTALMLPWGCDVQSLSLKHLHEEVLQVKGIL
 KNWRLTDEIRILGAFNRNPKNFQDNLNLRVSSDGMILAFMGVFNILQVLFDFR
 STKEFLPDPAVVRKLSQVLSKLTPEQERLRVGMQFDDDLISVITGAFKLVYA"
 BASE COUNT 934 a 696 c 748 g 841 t
 ORIGIN
 Query Match 94.0%; Score 2946; DB 9; Length 3219;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 3096; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 12 ggtggagaagcttccatccgaagaatgaattcacaagaacatagccgaagccctatt 71
 Db 121 GGTGGAGACCTTTCATCCCAAGCAATGAATTTTCATTAAGCAATAGCCGAGCCCTTATT 180
 QY 72 caaagaatggaatgactgtataaagaacaatcacagatgacctatttgatgaaatgt 131
 Db 181 CAAAGATGGAAATGACTGTATTAAAGCAATCACAGATGACCTATTGTATGAATGTT 240
 QY 132 ctgaatcgcgaaagaatgaacatcattgtctcgagaaggtgagcagaatgctgctaga 191
 Db 241 CTGAATCGCGAAGAAAGTAACATCATTTGCTCGAAGAGGTGAGAGGATGCTGCTAGA 300
 QY 192 gggatcattcacatgattttgaaaaaggctcagagctcgtgtaacctctttttaatcc 251
 Db 301 GGGATCATTCACATGATTTTGAANAAGGCTTCAGAGTCTGTAACTCTTTTCTTAATCC 360
 QY 252 ctlaagagatggaactatccctctattctcagaactgaaatggaacagcttcttccatca 311
 Db 361 CTTAAGAGATGGAATGACTATCTCTATTTCAGACTGAATGAGCAAAAGTCTTTTCATG 420
 QY 312 acatcaagaaggagacttggacgatttgctcagagattgaaggacttgcacatcccca 371
 Db 421 ACATCAGAAAGGAGACTTGGACGATTTGGCTCAGAGATTGAAGACTTGTACATACCCCA 480
 QY 372 tctttctgaacttataccctctgtgagaatgatgacattatttaacttgaagaagc 431
 Db 481 TCTTTCTGAATCTTTATATCCCTGTGTGAAGATATTGACATATTATTTAACCTGAAAAC 540
 QY 432 accctcacagaaccctgctctgtgagaagaacacacacatcacccgcttgagcagctg 491
 Db 541 ACCCTCACAGACCTGTCTGTGAGAGAAAGACCAACACATCACCGCGTGAGCAGCTG 600
 QY 492 accctgaatgagctcctgtgagagctcttcaagagccctgacatattgaaggaggaatctggc 551
 Db 601 ACCCTGAATGAGCTCCTGTGAGGCTCTTCAGAGCCCTGATCATTTGAAGGGAATCTGTGC 660
 QY 552 aaaggcaatcacctctgtctgagcagcattgacatgctctggggctcggaaagtgcag 611
 Db 661 AAAGCAATGCACCTGTGTGAGCAAGAAATGCAATGCTGTGGGCTCCGGAAGTGCAAG 720
 QY 612 gctctgaccaagttcaaatctctctctcctcgctcgaaggagccaggtgagactt 671
 Db 721 GCTCTGACCAAGTTCAAAATTCCTCTTCCTCCGCTTCAGCAGAGGCCAGAGGTGAGACT 780
 QY 672 ttggaacaccttgatgaactcctctgtatataactctgacacataaggaagaagcattc 731
 Db 781 TTTGGAACCTTGTGATCAACTCTCGATATACCTGGACATCAGCAATCAGAGCAATTC 840
 QY 732 atggccatgctctgaagctcgagcagagaggtcttcttctctctgagtgcaaatgaa 791
 Db 841 ATGGCCATGCTCTGAAGCTGCGGCAAGAGGTTCTTTTCTCTTGTATGCTATCAATGAA 900
 QY 792 ttcaagccccaagaactgcccagaanaatcgaaagccctgataaaggaacacacccgctcaag 851
 Db 901 TTCAGCCCCAGAACTGCCCAAGAAATCGAAGCCCTGATTAAGAGAAACACACCGCTTCAAG 960
 QY 852 aacatggtcatgtaaccactacatgagtgctcgaggcaaatcgaggttggtgctc 911

D	961	AACATGTCATGTGTCACCACTACCACTGAGTGGCTTGAGGCACATPACGGCAGTTTGGTGCC	1020
O	912	ctgactcgtcgaagtgaggagatalgacagaagaacgagccagagctcatalccagaagt	971
D	1021	ctgactgctgagtgaggatgatgacagaagacagccagagctctcatccagaagt	1080
O	972	ctgataaagagcttgctgaaggttgctgctcacaattcagaatccagtgctgaag	1031
D	1081	ctgattccaaagactctgctcaagccttgctccaattccagaattccagagctctgag	1140
O	1032	aattcattgaagaacccctcttgctgctcaactctgtgcaatccagaatgggtgaagt	1091
D	1141	aatctcatataaagccccctcttcttgctgctcattctgctcaattccagatgggtgaagt	1200
O	1092	gagttccactctcacaacacaacagcgtgtccatacctctcatgctgtgataag	1151
D	1201	gagttccattctcaccacacaacagcgtgtccatacctctcatgctgtgataag	1260
O	1152	aaaaaacaacacaataaagggtgctgctgaagtgaactcattcgaagcctgagccac	1211
D	1261	aaaaaacaacacaataaagggtgctgctgaagtgaactcattcgaagcctgagccac	1320
O	1212	tgtagagactgagctcctgtagaggtgtgtctcccaaaatttgatttgagctgcagat	1271
D	1321	tgtagagactgagctcctgtagaggtgtgtctcccaaaatttgatttgagctgcagat	1380
O	1272	gtgtcagcgttaatgtagaggtgtgtctcgtcgaacacgagcctcctgtataatacagt	1331
D	1381	gtgtcagcgttaatgtagaggtgtgtctcgtcgaacacgagcctcctgtataatacagt	1440
O	1332	caaaagttcaagccaaagataaataatctctcaaaagtcattccagaggtacacagcaga	1391
D	1441	caaaagttcaagccaaagataaataatctctcaaaagtcattccagaggtacacagcaga	1500
O	1392	cgaagactcaagcagtttatgtagctctcatgagccagaagaggtgaccaaagggaatgt	1451
D	1501	cgaagactcaagcagtttatgtagctctcatgagccagaagaggtgaccaaagggaatgt	1560
O	1452	tacttgcagaagaatggtttccatttcgagcatatacctaataagcagcctgcgcg	1511
D	1561	tacttgcagaagaatggtttccatttcgagcatatacctaataagcagcctgcgcg	1620
O	1512	tacacgtgtggtlcatctgtggaagccacagagcgtgttatgaaagcctcgcagcagtg	1571
D	1621	tacacgtgtggtlcatctgtggaagccacagagcgtgttatgaaagcctcgcagcagtg	1680
O	1572	tatcaaacagcgtcctctctgagacttccatccgccaagaagcctctctggaagacagaa	1631
D	1681	tatcaaacagcgtcctctctgagacttccatccgccaagaagcctctctggaagacagaa	1740
O	1632	tctttgcagaagctgtgaacaacacactgagcaagaatctgaaagccataaacatcaat	1691
D	1741	tctttgcagaagctgtgtaaaaacacactgagcaagaatctgaaagccataaacatcaat	1800
O	1692	tcccttgtagagtggtgcaatcatatatacaagaagatcatccaaatcagccctgagc	1751
D	1801	tcccttgtagagtggtgcaatcatatatacaagaagatcatccaaatcagccctgagc	1860
O	1752	caagaattggaagcttctcttcaaggtaaagcttatatacaactccagggaaacatcccc	1811
D	1861	caagaattggaagcttctcttcaaggtaaagcttatatacaactccagggaaacatcccc	1920
O	1812	gattactatgtagacttcttgacaattgccaactgtgcaagtgctcgtgagctcat	1871
D	1921	gattactatgtagacttcttgacaattgccaactgtgcaagtgctcgtgagctcat	1980
O	1872	aaactgagacttlataggggagctatgctcatalggaagaaagcgtcgaagaagcacagt	1931
D	1981	aaactgagacttlataggggagctatgctcatalggaagaaagcgtcgaagaagcacagt	2040
O	1932	ggaatccacatggaagagccccaagaacatcatccagcagagcgtatctgttc	1991
D	2041	ggaatccacatggaagagccccaagaacatcatccagcagagcgtatctgttc	2100

OY 3072 agtgtattacaggtgctttaactagtaactgcttaa 3110
 |||||||
 Db 3181 AGGTATTATACAGTGTCTTTAACTAGTAAGCTTAA 3219

RESULT 3
 LOCUS AY027787 3396 bp mRNA PRI 20-JUL-2001
 DEFINITION Homo sapiens CLAN1 (CLAN1) mRNA, complete cds.
 ACCESSION AY027787
 VERSION AY027787.1 GI:14324112
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
 AUTHORS Damiano, J.S., Stehlik, C., Pio, F., Godzik, A. and Reed, J.C.
 TITLE Clon, a novel human ced-4-like gene
 JOURNAL Genomics. 75 (1-3), 77-83 (2001)
 MEDLINE 21365712
 PUBMED 11472070
 REFERENCE 2 (bases 1 to 3396)
 AUTHORS Stehlik, C., Damiano, J.S., Pio, F., Godzik, A. and Reed, J.C.
 TITLE Direct Submission
 JOURNAL Submitted (21-FEB-2001) Program on Apoptosis and Cell Death
 Research, The Burnham Institute, 10901 North Torrey Pines Road, La
 Jolla, CA 92037, USA

FEATURES
 source location/Qualifiers

1..3396
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /map="2p22-p21"
 /tissue_type="lung"
 1..3396
 /gene="CLAN1"
 277..3351
 /gene="CLAN1"
 /codon_start=1
 /product="CLAN1"
 /protein_id="AAK14776.1"
 /db_xref="GI:14324113"
 /translation="MNFIKDNRSLIQRMGTVIKQITDLEFVNVNLRREVNIIICE
 KVEDAARGIIHMLIKGSESCNLFKLKIMNYPFDLNGOSLFEHSTSGDLDLA
 ODLKDYHTPFLNFEYPLGEDIIDIFNLSKPFETPEILRKPOHHRVQLTSLGLOA
 LOSPCITGESGSKSTLORIAMMGSGKALTRKEFVPELRSLRAOGLFETICD
 OLIDIPGIRKOTFMAMLLKROVYELLDYNEKPPONCEIEALIKENHRFKMYI
 VTTTECLRHTRFGALTAEGDMEDSAOLIREVLIKEALGELLQIORSCLRNLI
 MKTLPFVVITCAIQMGSESEFHSHTQTLTFHYDILLIQNRHKKHGAASDFIRSLDH
 RGDALBEVFSHKDFELQDVSYNEDELTLTGLCKYTAQRFKKYFEPKRSPOEYT
 AGRLSSLTSHPEEVTKNGYLOKMWISIDISTYSLLRYGSSVEADRAVMKH
 LAAYHSGCLIGLSIAKRLPLOSIOVKMTTOELIKATININSFVCGIHIQOEST
 SKMSLSQEFEPFGSKSLXINSNIPDLTFEPFELPSCASALDFIKIDPFGAMASH
 EKAADDTGGIIMEAPETYPISRAVSLEFNNKQERITLLEVLRLDSKINKODIRLKG
 IFSSATSLRIQIKCAGVAGSLSLSTCKNTYSLMWASPLTIEDERHITSVNLEKT
 LSIHDLQNRPLPGCLTDSLGNLKNLTKLIMNIMNEDAKIKLGLNLKMKCFLHL
 THLSDIGDMYIVKSLSEPCDLEIOLVSCCASNEDATILAOMLHVLKSLTILDS
 ENYLEKGSNEALHELIDRMNVLBDLTALPMGDDVGSLSILKHLEEVPOIKGL
 KNWRLTDEIRIILGAFPKNPLKFOOLNLAGNYSNGMLAPKQVFNKLKQLYPFDF
 STKEFLPDPALVRKLSOVLKSLTFLQEARLVGMQFDDDDLSVITGAFLVLA"

BASE COUNT 992 a 737 c 793 g 874 t
 ORIGIN

Query Match 93.0%; Score 2915; DB 9; Length 3396;
 Best Local Similarity 99.9%; Freq. No. 0;
 Matches 3115; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 12 ggtgggaagcttcacagaaatgaattcacaaggaacatagccgagcccttatt 71
 |||||||
 Db 253 GGTGGGAAGCTTTCATCCAGAACATGAAATTTCATTAAGACAAATAGCCGAGCCCTTATT 312

OY 72 caaagaatgggaatgactgtltaaaagcaatcacagatgacctatttgatgaatgtc 131
 |||||||
 Db 313 CAAGAATGGGAATGACTGTATTAAGCAATCACAGATGACCTATTGTATGGAATGTT 372
 OY 132 ctgaatcggaagaagaaatcaatcatcttgcgcgagaaggtggagcaagatgctctaga 191
 |||||||
 Db 373 CTGAATCGGGAAGAAAGTAACATCATTTGCTCGAAGAGGTGGAGAGATCTGCTAGA 432
 OY 192 gggatcatcatgatatttgaaaaagggttcagaagcttcgttaacctcttcttaaatcc 251
 |||||||
 Db 433 GGGATCATTCATGATGATTTTGAAGGAGGTTTGAAGAGTCTGTAACTCTTTTAAATCC 492
 OY 252 ctlaaggaatgggaatcatctatctatttcagaacttgaatgaaagtcctttcatcag 311
 |||||||
 Db 493 CTTAAGGAATGGAACTATCTCTATTTCAGGAGCTTGAATGGAAGTCTTTTATCATCAG 552
 OY 312 acatcagaagaagacttggagcaatctgtgcacagatttaaggaacttgcatacccca 371
 |||||||
 Db 553 ACATCAGAAGGAGACTTGGACGATTTGGCTCAGGATTTAAAGGACTTGACCATACCCCA 612
 OY 372 tctttctgaactttatccctctgtgagaatattgacattatttlaacttgaagaagc 431
 |||||||
 Db 613 TCTTTTCTGAATTTATATCCCTTGTTGTAAGATATTGACATTTATTTTAACTTGAAGAC 672
 OY 432 accctacagaacactgtctctgtgagaagaagcaaacacatcacgcgttggagagctg 491
 |||||||
 Db 673 ACCCTCACAGAACCTATCTGTGAGAGAGAACCAACACATCACGCGGTGAGCAGCTG 732
 OY 492 accctgaatgctctctgtgaagctcttcagaagccctgtgataatgaagaagaaatctgac 551
 |||||||
 Db 733 ACCCTGAATGGCTCTCTGAGGCTCTTCAGAGCCCTTGATATTGAAGGGAATCTGGC 792
 OY 552 aaagcaagctcacctctgtgcagcagatgtgcataatgttggggtccggaagtgcaag 611
 |||||||
 Db 793 AAAGCAATCAGCTCTGTCGACGACATGCTGCGGCTCCGGAAGTGCAG 852
 OY 612 gctctgacaaatgccaatctctctctcgcgtctcagaagggcccaaggtggaact 671
 |||||||
 Db 853 GCTGTGACCAAGTTCAAAATTCCTCTTCCTCCGCTCAGCAGGCGCCAGGCTGAGACTT 912
 OY 672 ttggaacccctgtgatatcaactcctgtgatatcctgtgacataatgaagaagacattc 731
 |||||||
 Db 913 TTGGAACCCCTGTGATATCACTCTGATATCACTGCGACATCAAGAGAGACATTC 972
 OY 732 atggcactctctcgaagctgcgcaagaggtctcttctctctgtatgtcacaaatga 791
 |||||||
 Db 973 AAGGCATCTCTCTGAAGCTGGGAGAGAGGTTCTTTCTCTTGATGCTACAAATGA 1032
 OY 792 ttcaagccccagaactgcccagaatcgaaagccctgataaaggaanaacacccgctcaag 851
 |||||||
 Db 1033 TTCAAGCCCCAGAACCTGCCAGAAATCGAAGCCCTGATTAAGGAAACACACCGCTTCAAG 1092
 OY 852 aacatggtcatgtatacacttccactgaatgagctgcgaagcaatacagaggttgctgc 911
 |||||||
 Db 1093 AACATGTCATGTCACCACTGACCACTGAGTCTCGAGGACATACGAGGATTTGGTGGCC 1152
 OY 912 cgtacgtcgtgaagtgaggataatgaagaagaacagcccgagctctcatcgaagaatg 971
 |||||||
 Db 1153 CTGACTGCTGAGAGTGGGAGATATGACAGAAGACAGCGCCAGGCTCTCATCGAAGAAG 1212
 OY 972 ctgatacagaagctgtcgaaggtgtgtgtccaaatccagaatccaggtgtcagg 1031
 |||||||
 Db 1213 CTGATCAAGGAGACTTGCTGAAGGCTTGTGCTCAAAATTCAGAAATCCAGTGTGAAG 1272
 OY 1032 aatctcatgaagaacccctctcttctgtgtcatcacttgtgcaatccagaatggatgaagt 1091
 |||||||
 Db 1273 AATCTCATGAAGAACCCCTCTCTTTGTGTGTCATCACTTGTCATCAATCCAGTGGGTAAAT 1332
 OY 1092 ggttcacactcacacacaaacaagcgtgtccatacctctcatgaltgtgtgataag 1151
 |||||||
 Db 1333 GAGTTCACACTGCACACAAACAAACAGCGTGTTCATCTTATGATCTGTGTATACAG 1392
 OY 1152 aaaaaaacaacacataaagggtgtgtgtcgaagtgaacttcatcgcgaagcttgagcaac 1211

Db 1393 AAAAACAACACAAACATTAAGGTGTGGTGCAGAGTACCTCATTTGGAGCCGAGACAC 1452
 Oy 1212 tgttgagacctgctctggagggtgtgtctccaccagaatttgaaattcgaaatcgagat 1271
 Db 1453 CCGTGGAGACCTGAGCTCTGGAGGGGTGTGTCTCCACAAAGTTGATTTGCAATGCAAGAT 1512
 Oy 1272 gttcagcgtgaatgaggatgtccctgctgacaactgggctcctcgttaatttaacgt 1331
 Db 1513 GTGTCCAGCGTAATGAGGATGTCTGCTGACACACTGGGCTCTCTGTAAATATACAGCT 1572
 Oy 1332 caaaggttcaagccaaagataaattctcttcaaaagtcataccaggagtaacagcagaa 1391
 Db 1573 CAAAGGTTCAAGCCAAAGATTAATCTTTCAAAAGTCAATTCAGAGATACACAGCAGGA 1632
 Oy 1392 cgaagacacagcagttatttgacgctcctcaatgacagagaggtgacccaagggaatgt 1451
 Db 1633 CCAAGACCTCAGAGGTTATTTGACGTCTCATGAGCCAGAGAGGTGACCAAGGGGAATGCT 1692
 Oy 1452 tacttgcaaaaatgtttccatttcgagacattacatccattatagcagggctgctccgg 1511
 Db 1693 TACTTGCAAAAATGGTTTCCATTTCCGACATTACATCCATTATAGCAGCCTGCTCCGG 1752
 Oy 1512 taacacctgtygtcactcgttggaagccacagggctgtatgaaagacacctcgagcagtg 1571
 Db 1753 TACACCTGTGGGTCAATCTGTGGAAGCCACAGGGCTGTTATGAAAGCACTCCGACGAGTG 1812
 Oy 1572 tatcaaacagcgtcctctcctcgacttccatccgccaagagcctctctggaagaa 1631
 Db 1813 TATCAACACGGGCTCTCTCGGACTTTCATCGCCAAAGGGCTCTCTGGAACAGGA 1872
 Oy 1632 tcttgcaaaagtgtgaaacacacacactgagcaagaaattctgaaagccataacatcaat 1691
 Db 1873 TCTTTGCAAAAGTGTAATAAACACCACTGAGCAAGAAATTTCTGAAGCCATAACATCAAT 1932
 Oy 1692 tcccttgtagagtgtygcatccattatatacaagagagatataccaaatcagccctgagc 1751
 Db 1933 TCCTTTGTAGAGTGTGGCATCATTTATATCAAGAGATACATCCAAATCAGCCCTGAGC 1992
 Oy 1752 caagaatttgaaagcttcttcaagtgaaagctatatacaactcagggagaaatcccc 1811
 Db 1993 CAAGAATTTGAAAGCTTTCTTTCAGGTAAACCTTATATATCAACACAGGAAACATCCCC 2052
 Oy 1812 gattactatttgacttcttgaacatttgccaatttgccaagtgccttgaaatcatt 1871
 Db 2053 GATTACTTATTTGACTTCTTTTAAACATTTGCCAATTTGCAAGTGCCTGGACTTCAT 2112
 Oy 1872 aaactggaactttaaagtgaggagctatgagctcaatggaagagcctgcaagaacacaggt 1931
 Db 2113 AAACGTGACTTTTATGGGGAGCTATGGCTTCATGGGAAAGGCTGCAGAAAGACACAGGT 2172
 Oy 1932 ggaatccacatggaagagggccccaagaaactacatctccagcagggcgtatcttgctc 1991
 Db 2173 GGAATTCACATGGAAGAGGCCCCAGAAACCTACATTCAGAGAGGGCTGTATCTTTGTTC 2232
 Oy 1992 ttcaactggaagcagaattcaggaactcgtgaggtgacacactccggagttcagcaagt 2051
 Db 2233 TTCAACTGGAAGCAGGAATTCAGAGACTCTGAGGTACACTCCGGGATTTTCAGCAAGTGG 2292
 Oy 2052 aataagcaagatalacatalcttgaggaaatalatcagctctgccaacagcctcagagctg 2111
 Db 2293 AATAAGCAAGATATCAGATATCTGGGGAAATATTCAGCTGCGCCAAAGCCGACAGGCTG 2252
 Oy 2112 caataaagaagatgtgcgtggtgtgctggaagccctcagtttgctcctcagcaactgt 2171
 Db 2353 CAAATTAAGAGATGTGCTGGTGTGGTGAAGACCTCAGTTGGTCTCAGCACTCTAAG 2412
 Oy 2172 aacattatctcctatgtgtgaaagcagttccctcacaatagaagaatagaagagcactc 2231
 Db 2413 AACATTATTTCTCATGTGTGAAGCAGATCCCTCACCATGAAATGAGAGGACATC 2472
 Oy 2232 acatctgtaacaaacctgaaacctgagatattacagactcagaaatcaagcgcgcg 2291

Db 2473 ACATCTGTAAACAAACCTGAAACCTTGAGTATTCATACCTACAGATTCACAGGCTCCG 2532
 Oy 2292 ggtgtctgactgaacagcttggtlaacttgaagaacctataaagctcataatgataac 2351
 Db 2533 GGTGTCTGACTGACACTTGCGGGTAACTGTGAAGAACCTTACAAAGCTCATATATGATTAAC 2592
 Oy 2352 ataaagaatgaagaagatgtcctataaactagctaaagcctgaanaacctgaagaag 2411
 Db 2593 ATTAAGATGAATGAGAGATGTCTATTAACCTTACTTAAGGCTGGAATAACCTGAAGAAG 2652
 Oy 2412 atgttatttcaatctggaacacactgtctgacacttggaagggatggaatgataatgct 2471
 Db 2653 ATGTATTATTTATTTATTTGACCACTTGTCTGACATTGGAAGGGGAATGATATATATGCT 2712
 Oy 2472 aagtcctctgaagtgaaacctgtgacacttgaagaatcaatagtcctcctgctgt 2511
 Db 2713 AAGTCTCTGCAAGTGAACCTGTGACCTTGAAACAAATTAATTAATGCTCTGCTGCTG 2772
 Oy 2532 tctgcaaatgcaatgaaatctcagctcagaatcttcaaatcttgatcgaactgagact 2591
 Db 2773 TCTGCAAAATCAGTGAATAATCTTAGCTCAGATCTTCAATTTGGTCAAACTGAGCATTT 2832
 Oy 2592 ctgatttatacagaataatctccttggaagaaatgaaatgaagcctctcactgaactgac 2651
 Db 2833 CTGATTTATTCAGAAATTAATCTGGAATAAGATGGAATGAAGCTTTCATGAAGCTGATC 2892
 Oy 2652 gacaagatgaagtgctcagaacagctcaacgcagctgactgtgctcctggtggtgtgagctg 2711
 Db 2893 GACAGATGAACGTGTCTGAGAACAGCTCACCGCATGTGTGCTGCGCTGGGCTGTGAGAGTG 2952
 Oy 2712 caagcagccttgagcaagcctggttgaaacatttgaagaggctcccaacactcgtcgaact 2771
 Db 2953 CAAGCGAGCTTGACAGACCTGTTGAAACATTTGGAGAGAGTCCCAACACTGTCAGACTT 3012
 Oy 2772 ggggtgaaaaaactggaagactcacagatcacagagatlaagaaatttaagtgtatcttctgga 2831
 Db 3013 GGGTTGAAAACTGAGACACTACAGATACAGAGATTGAATTTAGGTGCAATTTTTTGA 3072
 Oy 2832 aagaacctcttgaagaaacttccagcagctgaatttggcgggaaactcgttgaagcagtgat 2891
 Db 3073 AAGAACCTCTGAAAAAATTCACAGCACTTGAATTTGGCGGGAATCTGTGAGCAGAT 3132
 Oy 2892 ggtgtgcttgctcctcaatggtgtatcttgaagaacttgaagaatagagttttttgactt 2951
 Db 3133 GGTGCTTCTTCATGAGGTGTATTTGAGAACTTTAAGCAATTTAGTGTGTTTGTGACTTT 3192
 Oy 2952 agtactaaagaattctacactgataccagcatlaagtcagaanaacttaagcagatgatacc 3011
 Db 3193 AGTACTAAAGAAATTTGACTGATTCAGCATTTAGTGAAGAAACTTATACCAAGTGTATCC 3252
 Oy 3012 aagtttaacttcttgcagaagactaggtctgttgggtgccaatttgatgatgatctc 3071
 Db 3253 AAGTTAACTTTTGTGCAAGAGCTAGGCTGTGGGCGCAATTTGATGATGATGATCCTC 3312
 Oy 3072 aggttatttcaaggtgctttaaactagtaacagcttaaaagtgatctcgaagagca 3130
 Db 3313 AGTGTATTACAGGTGCTTTTAACTAGTAACTGTAAGTCTTAAATAAGTGTACTCCAGGCCA 3371

RESULT 4
 AF376061 AF376061 3581 bp mRNA PRI 15-MAY-2001
 LOCUS Homo sapiens caspase recruitment domain protein 12 mRNA, complete
 DEFINITION cds.
 ACCESSION AF376061
 VERSION AF376061.1 GI:14040074
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 3581)
 AUTHORS Gingras, M., Qiu, J. and Margolin, J. F.

TITLE Differential expression of the caspase recruitment domain protein 12 (CARD12) during monocytic differentiation

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 3581)

AUTHORS Gingras, M., Olu, J. and Margolin, J.F.

TITLE Direct Submission

JOURNAL Submitted (03-MAY-2001) Pediatric/Texas Children's Cancer Center, Baylor College of Medicine, 6621 Fannin St. MC3-3320, Houston, TX 77030, USA

FEATURES

source Location/Qualifiers

1..3581

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="2"

/map="2p21-22"

/cell_line="U937"

/cell_type="peripheral blood-derived CD14 mature monocyte"

490..3564

/note="CARD12"

/codon_start=1

/product="caspase recruitment domain protein 12"

/protein_id="AAK53443.1"

/db_xref="GI:1404075"

/translation="MNFIKDNRALIQRMGMVYIKQITDDLFVWVNLREPNITCCF
KVEDPAAGITIMILKSGESCONFLKSLKEMNPLFODLNGSLFHSBEDDLDA
ODLKDLYHTPSFLNYPLEGEDIDILFNKSTFEVLMRKQDHHRVQLTNGLLQA
LQSPCIEGESGKSKSTLQRIAMLGSGKALFEVFLRSRQGGFEFLCD
QLLDIPGTRKOTFAMMLKLRVLFLEFNEKPNCEIEALIKENHRKMYI
VTTTECRHRIROFALFAEYGDMEASQALIREVLKLEAGELIOLIOSRCRLN
MKTPLFVITCAIOGSESPHSHTQTLFHFYDILLKNNKHGVAASPIRSLDH
CGDLAEVGFHKRDEPDVSVNEDLITGLCKTAAORFKRYKPFHRSPEYT
AGRLSLLTSHPEEVTGNGYLDOKMVISDITSTVSLLRTGSSVEATRAVMKH
LAAVQHGCLGLSLAKRPLRMQESLQSVKMTTEQELIKAININSEVCGIHLQEST
SKSLSOFEFAFQKSLYINSNGNIDPLDFEFLPNCASLDPIKLDVFGGAAMW
EKAEDTGIHMEAEPIYIPSRVSLFENKQEFRTLEVLRLDFSKLNKODIRYLG
IFSSATSLRIQKRCAGVAGSLVLSCTKNYSLMVEASPLTIDEDEHITVTLKT
LSIHDLQORLPGLTSLGINKNTKILINDINKNEEDAIKLAGLKNKKMGCFHL
THLSDIGRMVYIKSLSSEPCDDEIDVLYSCCLSAVKILQNLHLVKSITDLS
ENYLEKDNENALHELIDRMNVLEDTALMLPWGCVQSSLSLKHLEVPQVLYGL
KNMRLTDEIRILGAFEGKNPLKNFOQLNLAGNRVSSDGLAFMGVFNPLQVYFDF
STKEFLPDPLVRKLSQVLSKITFLQEARLVGWOPDDDLVITGAFKLVYA"

misc_feature 502..741

/note="Region: caspase recruitment domain"

ASE COUNT 1033 a 781 c 843 g 924 t

RIGIN

Query Match 92.9% Score 2912; DB 9; Length 3581;
Best Local Similarity 99.9% Pred. No. 0;
Matches 3112; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 12 ggtgggaagcttcacacgaacatgaattcatataaagaacatagccgaagccctatt 71
DB 466 ggtgggaagcttcacacgaacatgaattcatataaagaacatagccgaagccctatt 525

QY 72 caaagaatgggaatgactgttataaagaacatcagaatgacctattgtatggaaatgt 131
DB 526 caaagaatgggaatgactgttataaagaacatcagaatgacctattgtatggaaatgt 585

QY 132 ctgaatcgcgaagaagtaaacatcatcttgctcgagaagaatggagaagaatcgtctata 191
DB 586 ctgaatcgcgaagaagtaaacatcatcttgctcgagaagaatggagaagaatcgtctata 645

QY 192 gggatcatcacatgatttgaaaaaagggttcagaatccctgaacatcctcttcaaatcc 251
DB 646 gggatcatcacatgatttgaaaaaagggttcagaatccctgaacatcctcttcaaatcc 705

QY 232 cttaaggagtggaactatccctctatttcaagaatgaaatgaaatgaaatgaaatgaaatg 311
DB 706 cttaaggagtggaactatccctctatttcaagaatgaaatgaaatgaaatgaaatgaaatg 765

QY 312 aacataagaagagacttgagacttggtcagattttaaagaactgtacacatcccca 371

DB 766 ACATCAGAGAGAGACTTGGACGATTTGGCTCAGAGATTAAAGACTTGTACATACCCCA 825

QY 372 tcttctgaactttatccctctgtggaagatattgacatatttttaacttgaagc 431

DB 826 tcttctgaactttatccctctgtggaagatattgacatatttttaacttgaagc 885

QY 432 accctcaagaacactgtctgtggaagaaagcaaacacatccacgcgttgagacgtg 491

DB 886 accctcaagaacactgtctgtggaagaaagcaaacacatccacgcgttgagacgtg 945

QY 492 accctgaatgactctctgaagcctctgaagccctctgaatattgaaggaatctgac 551

DB 946 accctgaatgactctctgaagcctctgaagccctctgaatattgaaggaatctgac 1005

QY 552 aaaggcaagctcacctgtctgacgcacatgcatgctctgagggtccggaagtgcag 611

DB 1006 aaaggcaagctcacctgtctgacgcacatgcatgctctgagggtccggaagtgcag 1065

QY 612 gctctgacaaatcgaatcgtctctctcgtcgtcgaaggccgaaggctgact 671

DB 1066 gctctgacaaatcgaatcgtctctctcgtcgtcgaaggccgaaggctgact 1125

QY 672 ttggaaccccttgatcaactcctctgatatcaactgacaaatgaagaagaatc 731

DB 1126 ttggaaccccttgatcaactcctctgatatcaactgacaaatgaagaagaatc 1185

QY 732 atgagcatgctgtgaagctcgagcagaaggtctcttctctctgactgactacaatga 791

DB 1186 atgagcatgctgtgaagctcgagcagaaggtctcttctctctgactgactacaatga 1245

QY 792 ttcagaagcccaagactgtcccaagaatcgaaagccctgataaagaagaacacgcgttcaag 851

DB 1246 ttcagaagcccaagactgtcccaagaatcgaaagccctgataaagaagaacacgcgttcaag 1305

QY 852 aacatgcatgctgtcaacactcaactcaactgagtcctggaagcaatacgaagctgtgac 911

DB 1306 aacatgcatgctgtcaacactcaactcaactgagtcctggaagcaatacgaagctgtgac 1365

QY 912 ctgactgctgaagtgagggaatatagaagaagaacagcccaagctctcaaccgaagtg 971

DB 1366 ctgactgctgaagtgagggaatatagaagaagaacagcccaagctctcaaccgaagtg 1425

QY 972 ctgatacaaggaactgtgtgaagctgtgtgtcccaaatccagaatccgaagctgtgag 1031

DB 1426 ctgatacaaggaactgtgtgaagctgtgtgtcccaaatccagaatccgaagctgtgag 1485

QY 1032 aatcctatgaagaccctctctgtgtgcatcaactgtgcaatccagaatgagtgaaagt 1091

DB 1486 aatcctatgaagaccctctctgtgtgcatcaactgtgcaatccagaatgagtgaaagt 1545

QY 1092 gattccactctcacacacacaaacgaagctgttccatlaaccttcatgactgttgaacag 1151

DB 1546 gattccactctcacacacacaaacgaagctgttccatlaaccttcatgactgttgaacag 1605

QY 1152 aaaaacaaac 1211

DB 1606 aaaaacaaac 1665

QY 1212 tgtgagacctagcctctgaggtgtgtgtcccaacaaagtgtgacttccaactgcaagat 1271

DB 1666 tgtgagacctagcctctgaggtgtgtgtcccaacaaagtgtgacttccaactgcaagat 1725

QY 1272 gttccagcgtgaaatgaagatgttccctgtgacaactgggtccctctgtaataatacagct 1331

DB 1726 gttccagcgtgaaatgaagatgttccctgtgacaactgggtccctctgtaataatacagct 1785

QY 1332 caaaggttcaagccaaagtataaatcttccaagatcattccagaggtatacagaagga 1391

DB 1786 caaaggttcaagccaaagtataaatcttccaagatcattccagaggtatacagaagga 1845

QY 1392 gaaagactcaagcattatgacgtctcaatgagcagaagaggttgaaccaaggaatggt 1451

DB 1846 gaaagactcaagcattatgacgtctcaatgagcagaagaggttgaaccaaggaatggt 1905


```

|||||
Db 50148 TCGAAGCTGAGAGTGTCCAGGCTGATAGATGCTCTGCTGCAACTGGGCTCCCT 50207
OY 1318 gtaaatatagatcaaaagttcaagccaaafataatctttcaagttcaatccagg 1377
Db 50208 GTAAATATACACTCAAGGTTCAAGCCAAAGTAAATTTCTTCAAGTATTCAGG 50267
OY 1378 agtaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1437
Db 50268 AGTACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 50327
OY 1438 ccaagggaggaatggttactctgcagaaatggtttccattccgagcattacatcattata 1497
Db 50328 CCAAGGGAATGCTTACTTGCAGAAATGCTTTCATTTCAGCAGTATTCATCCTTA 50387
OY 1498 gcaagcctgctccggtacacctgtggtgcatctgtggaagcacaagggcgtgtatgaac 1557
Db 50388 GAGGCTCTCCGCTACACTGTGGGTCACTGTGAGAACCCAGGGGCTGTATGAGAC 50447
OY 1558 accctgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1617
Db 50448 ACCCTGCAGCAGTGTATACAGCGCTGCTTCGACCTTCATCAGCAGCAGGCTC 50507
OY 1618 tctggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1677
Db 50508 TCTGGAGACAGGAACTTTCGAAAGTGAAGAAACACACAGCAGCAAGAAATCTGAAG 50567
OY 1678 ccataaacatcaatctctctctctctctctctctctctctctctctctctctctct 1737
Db 50568 CCATAAACATCAATCTCTTGTGAGAGTGGCATCATCTATTCAGAGAGTACATCCA 50627
OY 1738 aatcagccctgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1797
Db 50628 ATTCAGCCCTGAGCAGCAAGATTTGAAGCTTTCTTCAAGATTAAGCTTATATATCA 50867
OY 1798 cagggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 50927
Db 50688 CAGGAGACATCCCGATTAATTTGACCTTTTGAACCTTTCAGCAATTCAGAGAG 50867
OY 1858 ctctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 50807
Db 50748 CTCTGAGCTTCATTAACCTGAGCTTTTATGGGAGAGCTATGGCTTCAATGGAGAA 1977
OY 1918 cagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 50867
Db 50808 CAGAGACACAGCTGAGATTCACATGAGAGAGGCCCCAGAAACCTACATTCACAGAG 50867
OY 1978 ctgtatcttctctctctctctctctctctctctctctctctctctctctctctct 50927
Db 50868 CTGTATCTTCTTCTCACTGAGAGAGAGAAATTCAGAGCTGTGAGAGTCACTCCGG 2097
OY 2038 attcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 50987
Db 50928 ATTTCAGCAAGTGTATTAAGCAAGATATCTGAGAGAGAAATTTACACTCTGCGCA 50987
OY 2098 caagcctgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 51047
Db 50988 CAAAGCTCAAGCTGCAAAATTAAGAGATGTCTGCTGTGAGAGCTCTAGTTGTGCTC 51047
OY 2158 tcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2217
Db 51048 TCACGACCTGTAAAGCAATTTATCTCTCATGCTGAGAGCCAGCCCTCACATGAGAG 51107
OY 2218 atgagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2277
Db 51108 ATGAGAGGCAATCATCTGTAACAAACCTGAAACCTGTGATTCATACCTACAGAG 51167
OY 2278 atcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2294
Db 51168 ATCAAGGCTGCGGGGT 51184

```

RESULT 6

```

CNS01D53/c 138909 bp DNA PRI 02-MAR-2000
LOCUS 138909 bp DNA PRI 02-MAR-2000
DEFINITION BAC sequence from the SPG4 candidate region at 2p21-2p22 BAC 164M19
of C17B_978_SKB library from chromosome 2 of Homo sapiens (Human),
complete sequence.
ACCESSION AL121653
VERSION 1
KEYWORDS HTG; HTGS_DRAFT; SPG4 genomic DNA interval.
SOURCE human
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 138909)
AUTHORS Hazan, J., Fomnechten, N., Mavel, D., Paternotte, C., Sanson, D.,
Artiguenave, F., Davoine, C., Dur, A., Wincker, P.,
Bottier, P., Catillo, J., Barbe, V., Burgunder, J.M.,
Prid'Homme, J.F., Bricle, A., Fontaine, B., Heilig, R. and
Weissenbach, J.
Spastin, a novel AAA protein, is altered in the most frequent form
of autosomal dominant spastic paraplegia
Nat. Genet. (1999) in press
2 (bases 1 to 138909)
JOURNAL Genome.
REFERENCE Direct Submission
AUTHORS Submitted (29-FEB-2000) to the EMBL/Genbank/DBJ databases
JOURNAL Submitted (29-FEB-2000) to the EMBL/Genbank/DBJ databases
COMMENT On Mar 6, 2000 this sequence version replaced g1:602386.
FEATURES
location/Qualifiers
Source
1..138909
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/clone_lib="C17B_978_SKB"
/clone="164M19"
BASE COUNT 39243 a 28424 c 29121 g 42121 t
ORIGIN
Query Match 58.9% Score 1844; DB 9; Length 138909;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1994; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 298 gtcttttcatagacatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 75250
Db 75309 GTCTTTTCACTACAGATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 417
OY 358 tgcacatcacccatctctctctctctctctctctctctctctctctctctctctctctctctct 75190
Db 75249 TGATCAATACCCATCTTCTGACTTTTANCCCTTGAGAGATTAATGACATTAATT 75190
OY 418 ttaacttgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 75130
Db 75189 TTAACTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 537
OY 478 gctgtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 75070
Db 75129 GCGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 597
OY 538 aagggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 75010
Db 75069 AAGGGGAATCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 657
OY 598 ccggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 74950
Db 75009 CCGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 717
OY 658 cccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 74890
Db 74949 CCCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 777
OY 718 ggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 74830
Db 74889 GGAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG

```

OY	778	atcgctacaaatgaaatcctcaagcccaagacgtgcgcagaataatcgaaagccctgataaaggaaa	837
Db	74829	atggctacatataattccaagcccccgaacgtgccccaatctcaaacccctgataaaggaaa	74770
OY	838	accaccgccttcaagacatggtcatcgttcaaccaatacaactgtagtgcctgagacatac	897
Db	74769	accaccgcgttcaagaacatggctatgctacaccatgacactgtagtgcctgagacatac	74710
OY	899	ggcagtttggtgcccctgactgctgagtggtgggatagaacagaagaagcgccagctc	957
Db	74709	ggcagtttggtgcccctgactgctgagtggtgggatagaacagaagaagcgccagctc	74650
OY	958	tcaatccgagaagtgctgatacaagagagcttgctgaaagccttgctgccaatctgaat	1017
Db	74649	tcattccgaagactgctgatacaagagagcttgctgaaagccttgctgccaatctgaat	74590
OY	1018	ccaagtgcttgagaatcctcgtaaagacccctcctcttggtgtaactgtagaatcc	1077
Db	74589	ccaagtgcttgagaatcctcgtaaagacccctcctcttggtgtaactgtagaatcc	74530
OY	1078	agaatgggtgaaagtgaattccactctcaacaaacaaacgcgtgctacatcctatg	1137
Db	74529	agaatgggtgaaagtgaattccactctcaacaaacaaacgcgtgctacatcctatg	74470
OY	1138	atccttgaataagaaaaaaacaaacaaacataaagggtgtagctcgaatgtaattac	1197
Db	74469	atccttgaataagaaaaaaacaaacaaacataaagggtgtagctcgaatgtaattac	74410
OY	1198	ggaagcctgaaacactgtggaagacactgctggaaggtgtgtctcccaaatgtgat	1257
Db	74409	ggaagcctgaaacactgtggaagacactgctggaaggtgtgtctcccaaatgtgat	74350
OY	1258	tcgaaactgcaggaatggtgtccagcgtgaaatgagaatgtccctgtaacactggctcct	1317
Db	74349	tcgaaactgcaggaatggtgtccagcgtgaaatgagaatgtccctgtaacactggctcct	74290
OY	1318	gttaaatataaagctaaaggttcaagccaagtataaatctcttcaagaatgtaacag	1377
Db	74289	gttaaatataaagctaaaggttcaagccaagtataaatctcttcaagaatgtaacag	74230
OY	1378	agtaacagcagaagaaactcagaactcagatltatagctctcatagcagaagaggtga	1437
Db	74229	agtaacagcagaagaaactcagaactcagatltatagctctcatagcagaagaggtga	74170
OY	1438	ccaaggggaatggttactgtcagaataatggtttccattctcgacaataatccattata	1497
Db	74169	ccaaggggaatggttactgtcagaataatggtttccattctcgacaataatccattata	74110
OY	1498	ggaagcctgcgcgtaaacctgtgttgatctgttgaaagccaacgaagcgtgtatgaagc	1557
Db	74109	ggaagcctgcgcgtaaacctgtgttgatctgttgaaagccaacgaagcgtgtatgaagc	74050
OY	1558	acctgcagcaatgfatcaacaacagcgtgcctctcgcgaacttcaatgcagaagagctc	1617
Db	74049	acctgcagcaatgfatcaacaacagcgtgcctctcgcgaacttcaatgcagaagagctc	73990
OY	1618	tcctgagaaggaatccttgcaaaagtgtgaaanaacacacacttgagcaagaatcttgaaag	1677
Db	73989	tcctgagaaggaatccttgcaaaagtgtgaaanaacacacacttgagcaagaatcttgaaag	73930
OY	1678	ccataaataatcattcctcttgtagaggtgtgtgcatcattatataaagaagataatcca	1737
Db	73929	ccataaataatcattcctcttgtagaggtgtgtgcatcattatataaagaagataatcca	73870
OY	1738	aatcagccctgagcaagaatctgtagagcttcttcaaggttaaaagctatatacaact	1797
Db	73869	aatcagccctgagcaagaatctgtagagcttcttcaaggttaaaagctatatacaact	73810
OY	1798	caaggagaatccccgattactattatgactcttgaacaattggccaattgtgcaatg	1857
Db	73809	caaggagaatccccgattactattatgactcttgaacaattggccaattgtgcaatg	73750
OY	1858	ctctggaactcaataactggaactttatgtgggagataatgctcatatggtgaaaagcgtg	1917

Db	73749	CCCTGGACCTCATTAACCTGGACCTTTATGTGGGGAGCTATGCGCTTCATGGGAAAGGCTG	73690
QY	1918	cagaagacacagtggaatccaatggaagagccccaagaacataatcccaagcagg	1977
Db	73689	CAGAGGACACAGGTGGAAATCCACATGGAAGAGGCCCAAGAAACCTCAATTCACAGAGGG	73630
QY	1978	ctgtatcttgttcttcaacttggaagcaggaatlcaggaactctggagtgcaactccgg	2037
Db	73629	CTGTATCTTGTGTTCTTCACTGGAGCGAGGAATTCAGGACTCTGGAGGTCACACTCCGGG	73570
QY	2038	attccagaagtgaataagaagaatcacaatctacatctcggggaaataatccagctcgca	2097
Db	73569	ATTTCAGCAAGTTGAATAAGCAGAGATTCAGATCTGGGGGAAAAATTCAGCTCGCCA	73510
QY	2098	caagcccaagcctcacaataaagatgctgtgtgtgtgaagcccaatttgctcc	2157
Db	73509	CAAGCTCAGGCTCCTCAAAATTAAGAGATGCTGCTGTGGCTGGAAACCTCAGTTTGCTCC	73450
QY	2158	tcaagcactgtaagaacattatctctcatalgtgtggaagccagtcgccctacacataaag	2217
Db	73449	TCAGCACCCTGTAAAGAACATTTATTTCTCTCATGTGTGGAAGCAGTCCCTCACCATTAAGG	73390
QY	2218	atgaagagcacatcaatctgtaacaaacctgaaacctgaagtaatcaatgaactacaga	2277
Db	73389	ATGAGAGGACATCATCATCTGTAAACAACCTGAAACCTTGAGTATTTCATGACACTCACA	73330
QY	2278	atcaagagctgcggggt	2294
Db	73329	ATCAAGGCTGCCGGGT	73313

RESULT	7
IR2005417	
LOCUS	IR2005417
DEFINITION	Homo sapiens mRNA full length insert cDNA clone EUROLIMAGE 2005417
ACCESSION	AL389934
VERSION	FLJ8934.1
KEYWORDS	GI.9367839
SOURCE	FLI_CDNA.
ORGANISM	human. Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 1355) Auffray,C., Ansorge,W., Ballabio,A., Estivill,X., Gibson,K., Lehrach,H., Poustka,A. and Lundeberg,J. The European IMAGE consortium for integrated Molecular analysis o human gene transcripts Unpublished 2 (bases 1 to 1355) Pluviner,R., Estivill,X., Escarceller,M. and Sumoy,L. Direct Submission Submitted (15-JUL-2000) Dept. Genetica Molecular, Institut de Recerca Oncologica (IRO), Hospital duran i Reynals, Av. Gran Via s/n Km 2,7 L'Hospitalet del Llobregat, 08907 Barcelona, Catalunya, SPAIN. Tel.: +34-93-260-7775 Fax: +34-93-260-7776 WWW site: http://www.iro.es e-mail enquiries: Isumoy@iro.es EURO-IMAGE Consortium Contact: Auffray C
COMMENT	

Tel: ++33-1-49 58 34 98
 Fax: ++33-1-49 58 35 09
 e-mail: auffray@infobiogen.fr
 This clone is available royalty-free through IMAGE Consortium Distributors.
IMPORTANT: This sequence represents the full insert of this IMAGE cDNA clone. No attempt has been made to verify whether this corresponds to the full-length of the original mRNA from which it was derived.
 location/Qualifiers
 1. .135
 /organism="Homo sapiens"


```

/db.xref="taxon:9606"
/chromosome="2"
/clone="IMAGE cDNA clone 2005417"
/clone_11b="NCI_CGAP_Pan1"
/map="2p21-p22"
<1..1321
/codon.start=2
/product="hypothetical protein, weakly similar to
(AF102871) neuronal apoptosis inhibitory protein 2 [Mus
musculus]"
/protein_id="CAB97523.1"
/db.xref="GI:9367840"
/translation="INSGNIPDPYLDPEPEHLPCNASALDFTKLDPYGAMASWEKAE
DTGSIHMEAEPTIYIPSRVSLFFPNWKOEFTLEVTLDPSKLKQDRIYGLTFSSR
TSLRQIKRCGAVAGSLVLTSTCRNYSILWAEVSPITIDERRHITSTYNTKLTISID
LQNRDPGGTLSDLGNTLKNLKLIMDNIKNMEEDAIKLABGLNMLKMCLEHLHLSD
IGEMRDYVKSLSSEPCDLIEIQLVSCSLANAVKILAOINLHNLVYKLSIIDLSENYLE
KDGEALHELIDRNNVLEOLPALMLPMGCDVQGSLSLKLHEVPOLYKGLKNWRIR
TDFETRIILGAEFFGNPLKLPOLMLAGNVSDDCMIAFMCVFNELKOLYFFDPSTKEEF
LPDPAIVRKLSQVLSKLTFLQEARLVGQFDDDLVITTGAFKLVTA"

```

	Query Match	41.2%	Score 1290;	DB 9;	Length 1355;	
	Best Local Similarity	99.9%;	Prod. No. 0;	Mismatches 1;	Indels 0;	Gaps 0;
	Matches 1340;	Conservative 1;	Overlapped 0;	Indels 1;	Indels 0;	Gaps 0;
OY	1790	tatcaactcaaggagacatcccccattactatttactctcttgaacatttgccaatlg	1849			
Db	1	TATCAACTCAGGAGACATCCCGATTACTTATTACTCTTTGAAACATTGCCAATTG	60			
OY	1850	tgcgaagtctctggaactcaatcaactggaactttatggygagactatgcttcaatgg	1909			
Db	61	TGCAGTCTCTGGAAGTCAATTAACCTGACTTTATGGGGAGCTATGGCTTCATGGGA	120			
OY	1910	aaagctctcagaagaacacagtgtygaatccacatgtaaaagagccccaagaactaatcc	1969			
Db	121	AAAGCTCTCAGAAAGACACAGGTGGATCCACTGGAAAGGCCCAAGAACTCAATTCC	180			
OY	1970	cagcagggctgtatcttcttcttcaactggaagcagaatcagaactctggaagtcac	2029			
Db	181	CAGCAGGGCTGTATCTTTTGTCTTCAACTGGAAGCAGGAATTCAGGACTCTGGAGGTAC	240			
OY	2030	actccgggaatttcagcaagttgaaatgaagaacataatccatccctgggnaaataatcag	2089			
Db	241	ACTCCGGGATTTACACCAAGTTGAATTAACAGATATCTCGGGGAAATATTTCAG	300			
OY	2090	ctctgcacaagcctcagggctgcaaaataaagaatgtyctggtgtgcttgaagcctcag	2149			
Db	301	CTCTCCCAACAGCCCTCAGGCTGCAATTAAGAGATGCTGTGCTGGCTGGAAGCCTCAG	360			
OY	2150	tttgtctcagcagctgttaagaacattatctctcctgtyggaagcaggtccctcac	2209			
Db	361	TTTGTCTCTCAGCACCTTGAAGAACATTTATTTCTCTCATGTGGGAAGCAGTCCCTCAC	420			
OY	2210	catagaagaatgagagacacatacatctctgtaacaacacttgaacacttgatattcaga	2269			
Db	421	CATAGAAGATGAGAGACATCATCTCTGTAAACAACCTGGAATTCATGGA	480			
OY	2270	cctacagaatcaacggctgcgcgggtgtgtctgactgacagcttgygtaacttgaagaacct	2329			
Db	481	CCTACAGATCAACAGCGCTGCCGGGTGCTGTGACTGACAGCTTGCGTACTTGAAGAACCT	540			
OY	2330	tacaaagctcataatgagataacataaagaatgaatgaagaagaatgctataaaactagctga	2389			
Db	541	TACAAAGCTCATTAATGATTAACATTAAGATGATGAAGAGATGCTATTAATAACTAGCTGA	600			
OY	2390	aggccttgaaaaaaccttgaagaagaatgtgttatttcaatttgcccacttgcttgcacattgg	2449			
Db	601	AGGCTTGAAAAACCTGAAGAAGATGTGTTTATTATTATTTGACCCACTTGTCTGCACATTGG	660			
OY	2450	agaaggaaatgatatcatatagtaagctctctgtcgaatgaacacttgtaacttgaagaat	2509			

Db	661	AGAGGAAATGAGTATACATATGTCAACTCTCTGTGCAAGTAAACCCCTGTGACCTTTGAGAAAT	720
Qy	2510	tcaatagctctccgtcgtctbctgcgcaatbcaq tgaanaatcctagctcaaatcttca	2569
Db	721	TCAATTAGCTCTCTCTGTTGTCTGTCGCAAAATSCATGAAATTCGATGCTGAAATCTTCA	780
Qy	2570	caattbglcgaactcgaacatctctgattlalcagaanaatactcctbgaanaaatbga	2629
Db	781	CAATTTGGTCAACAGACGACATCTTGATTATTCAGAAATTAACCGGAAAAAGATGAA	840
Qy	2630	tbaaactctctcagacgatctgacaagatbgaagtgctagaanaagctcaccgcacgat	2689
Db	841	TBAACCTCTTCAGAACCTGATCAGACAGATGAACTGTGTAAACAGCTCACCGACATGAT	900
Qy	2690	gctgcgcctbgaagctbgaagtcgaagcagcctcgaagcctgttgaanaattcgaaga	2749
Db	901	GCTGCCCTGGGGCTGTGACGTGCAAGGCAAGCTGTAGCAGCCTGTTGAACATTTGGAGA	960
Qy	2750	ggtcccaacaactcgtcgaagctbgaagaaactbgaanaactcgaagctcagaatagatbga	2809
Db	961	GGTCCCAACAACCTGCAAGCTGTGGTGAATAAACGTGAGACTCACGATACAGAGATTAG	1020
Qy	2810	aatttagtgatatttttggaaagaaccctctgaanaacttccagcagttgtaattggc	2869
Db	1021	AATTTTAGTGATTTTTTGGAAACAAACCCCTGCAAAAACCTTCAGCGTTGAATTTGGC	1080
Qy	2870	gggaactcgtcgaagcaatgatatbgaatgcttgccttcaaggcttgaattggaactctaa	2929
Db	1081	GGGAATCGTGTGACACATGATGATGGCTTGCCCTCAATGGGTGTATTTGAAATCTTAA	1140
Qy	2930	gcaattagtgltttttgacttagtactaagaatcttccactgataccagcatagtcag	2989
Db	1141	GCAATTAGTGTTTTTGGACTTTAGTACCTTAAGAAATTTCTACTCATCCAGCATTAGTCAG	1200
Qy	2990	aaaacttgcgaagtgatataccaagttaacttctcgaagaacatagccttgttgggtg	3049
Db	1201	AAAACCTTACCCAGTGTTATCCAAATTAACCTTTTCTGCAAAAGCTAGGCTTGTTGGGTG	1260
Qy	3050	gcaatttgaatgatactcgaagtgatatacagtgatatacagtgctttaaactagtaactgtta	3109
Db	1261	GCAATTTGATGATGATGATCAGTCAGTGTTATTATACAGGCTCTTTAAACTAGTAAGTGCCTA	1320
Qy	3110	aataaagtgatctgaaagcca 3130	
Db	1321	AATAAAGTGATCTCGAAGCCA 1341	

RESULT	8
AY027788	
LOCUS	
DEFINITION	AY027788 Homo sapiens CLABD (CLAN1) mRNA
ACCESSION	AY027788
VERSION	AY027788.1 GI:14324114
KEYWORDS	
SOURCE	
ORGANISM	human.
REFERENCE	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 1395)
TITLE	Damiano,J.S., Stehlik,C., Pio,F., Godzik,A. and Reed,J.C.
JOURNAL	Clan, a novel human ced-4-like gene
MEDLINE	Genomics. 75 (1-3), 77-83 (2001)
PUBMED	21365712
REFERENCE	11472070
AUTHORS	2 (bases 1 to 1395)
TITLE	Stehlik,C., Damiano,J.S., Pio,F., Godzik,A. and Reed,J.C.
JOURNAL	Direct Submission
FEATURES	Submitted (21-FEB-2001) Program on Apoptosis and Cell Death
source	Research, The Burnham Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA
	Location/Qualifiers
	1..1395

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2p22-p21"
/tissue_type="Lung"
1..1395
/gene="CLAN1"
277..1356
/gene="CLAN1"
/codon_start=1
/product="CLANB"
/protein_id="AAK14777.1"
/db_xref="GI:14324115"
/translation="MNFIKDNRSLIORMGMTVYKQITDDELFWNVNLRREVNTICE
KVEDDAAGIITHMLIKGSESCNLFKSLKEMNYPLEFDLNGOSLFDLSLGMKMLTK
LIMDNIKMNEEDALILAGLKNLKMCLFHLTHLIDEGMDYIVKSLSEPCDLEET
OLVSCCLSANAVKILAIQNLHNLVKLSLIDSENLYEKDGNELHELIDRMNVEQLTA
LMLPMGCDVOSLSLKLKEEVPOLVATGKLMWLTDEIRIILGAFGKNLKNLFOO
LNLGNRYSOSGMLAFMGVFNELKQIVFDFSTKREFLDPLALVRKLSOVLKLTFLQE
ARLVGMQDDEDDLSITGAFKLVTA"
BASE COUNT      436 a      248 c      327 g      384 t
ORIGIN

Query Match      26.7%: Score 836; DB 9; Length 1395;
Best Local Similarity 100.0%: Pred. No. 0;
Matches 836; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2293 gtcgtctgactgacagctgtggtaacttgaagaaccttaagaagctcataatgataca 2352
|||||
DB 539 GTGGTCTGACTGACAGCTGGGTAACCTTGAAGACCTTCAAAAGCTCAATATGATAACA 598

OY 2353 taaagatgaatgaagaatgtctataaaactagctgaagcgctgaaaaacctgaagaaga 2412
|||||
DB 599 TAAAGATGAATGAAGAAGATGCTATTAACCTAGCTGAAGCGCTGAAAAACCTGAAGAAGA 658

OY 2413 tctgttattatcatcttgaccacctgtctgacatitgagaagggaatgataatagta 2472
|||||
DB 659 TGTGTTATTATTATTTGACCCACTTGTCTGACATTGAGAGGGAATGATTCATATAGTGA 718

OY 2473 agctctgtcaagtgaaacctgtgacctgaagaatltcaatagctctgtctgt 2532
|||||
DB 719 ACTCTGTCGACAGTGAACCCCTGACCTTGAAGAATTCATATAGCTCTGCTGCTTGT 778

Y 2533 ctgcaaatgcaatgaaaaatccctagctcagaatcctcaaatgtgtcaaatcgacattc 2592
|||||
DB 779 CTGCAAAATGCAATGAAAAATCTAGCTCAGAAATCTTCACAAATTTGGCAAACTGACATTC 838

OY 2593 ttgattatcagaataattacctggaagaagaatgaaatgaagctctctcaatgacatgacg 2652
|||||
DB 839 TTGATTTATCAGAAAATTTACCTGGAAGAAAAGATGAAATGAAGCTCTTTCATGAACGTGATCG 898

OY 2653 acaagatgaacgtgtctagaacacgtcaaccgacatgctgcctcgaggcgctgtgaagtcg 2712
|||||
DB 899 ACAGATGAACTGCTAGAACAGCTCACCGCAGCTGATGCTGCCCTGCGGCGCTGACGCTGC 958

OY 2713 aaggaagcctggaagcctgttgaacaatttggaaggaaggtcccaaacctgctcaagcttg 2772
|||||
DB 959 AAGGACGCTGAGCAGCTGTTGAAACATTTGGAGAGAGCTCCACAACTGCTCAAGCTTG 1018

OY 2773 ggttaaaaaacggaagctcacagatacagagattagaattttagttagattttttgga 2832
|||||
DB 1019 GGTAAAAAACGAGACTCACAGATACAGAGATTAGAAATTTTAGGTGCTATTTTGGGA 1078

OY 2833 agaaccctctgaaaaaactccagagcttgaatttggcggaagaatcgltgagcaatgacg 2892
|||||
DB 1079 AGAACCCTCTGAAAAAATTCAGAGTGAATTTGGCGGGAATGCTGTGACAGATGAGG 1138

OY 2893 gatgctgtgctcaatgggtgtatttgagaatccttaagcaatlagtlttttgaactta 2952
|||||
DB 1139 GATGCTGTGCTTCAATGAGTGTATTTGAGAAATCTTAACCAATTAAGTGTGTTTGA 1198

OY 2953 gtactcaagaatcttactctgatacagcatagtcagaagaatctagccaagtgatcatcca 3012

```

```

|||||
DB 1199 GTCATAAAGAAATTTCTACCTGATCCAGCATTTAGTACAGAAAACCTTACCAAGCTTATCA 1258

OY 3013 agttaaccttctgcaagaagcctagagctgtgtgggtggaatttgatgtatcata 3072
|||||
DB 1259 AGTTAACTTTTCTGCAAGAGAGCTAGGCTGTGTGGCGCAATTTGATGATGATCTCA 1318

OY 3073 ggttattacaggtgtctttaactagtaactgctcaatgaatgaatgactgaagc 3128
|||||
DB 1319 GTCTTATATACAGCTGCTTTTAAACGTAGTACGCTTAAATTAAGTGTACTCGAAGC 1374

RESULT 9
AY027789 768 bp mRNA PRI 20-JUL-2001
LOCUS Homo sapiens CLAN1 (CLAN1) mRNA, complete cds.
DEFINITION
ACCESSION AY027789
VERSION AY027789.1 GI:14324116
KEYWORDS
SOURCE
ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 768)
AUTHORS Damiano,J.S., Stehlik,C., Pio,F., Godzik,A. and Reed,J.C.
TITLE Clan, a novel human ced-4-like gene
JOURNAL Genomics. 75 (1-3), 77-83 (2001)
MEDLINE 21365712
PUBMED 11472070
REFERENCE 2 (bases 1 to 768)
AUTHORS Stehlik,C., Damiano,J.S., Pio,F., Godzik,A. and Reed,J.C.
TITLE Direct Submission
JOURNAL Submitted (21-FEB-2001) Program on Apoptosis and Cell Death
Research, The Burnham Institute, 10901 North Torrey Pines Road, La
Jolla, CA 92037, USA

FEATURES
source location/Qualifiers
1..768
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2p22-p21"
/tissue_type="Lung"
1..768
/gene="CLAN1"
/gene="CLAN1"
277..747
/codon_start=1
/product="CLANB"
/protein_id="AAK14777.1"
/db_xref="GI:14324115"
/translation="MNFIKDNRSLIORMGMTVYKQITDDELFWNVNLRREVNTICE
KVEDDAAGIITHMLIKGSESCNLFKSLKEMNYPLEFDLNGOSLFDLSLGMKMLTK
LIMDNIKMNEEDALILAGLKNLKMCLFHLTHLIDEGMDYIVKSLSEPCDLEET
OLVSCCLSANAVKILAIQNLHNLVKLSLIDSENLYEKDGNELHELIDRMNVEQLTA
LMLPMGCDVOSLSLKLKEEVPOLVATGKLMWLTDEIRIILGAFGKNLKNLFOO
LNLGNRYSOSGMLAFMGVFNELKQIVFDFSTKREFLDPLALVRKLSOVLKLTFLQE
ARLVGMQDDEDDLSITGAFKLVTA"
BASE COUNT 218 a 157 c 180 g 213 t
ORIGIN

Query Match 15.5%: Score 485; DB 9; Length 768;
Best Local Similarity 100.0%: Pred. No. 97e-251;
Matches 485; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 ggtgggaagccttcatccagaacaatgaattcataaaggaacaatagccagcccttatt 71
|||||
DB 253 GGTGGGAAGCTTTCATCCGAAACAATGATTTCAATAAGGACAAATAGCCGACCTTATT 312

OY 72 caaagaatgggaatgacgttatataagcaaatcaagaatgacctattgtatggaatgt 131
|||||
DB 313 CAAGAATGGGAATGACGTATATAAGCAAAATCAAGATGACCTATTTGTATGAAATGT 372

OY 112 ctgaatcggaagaatgaacatcatcttgcgcgagaaggtgagacagatgctctaga 191
|||||
DB 373 CTGAATCCGGAAGAAGTAAACATCATTTTGCTGCGAAGAGGTGAGCAGGATGCTCTAGA 432

```


OY 192 gggatcattacatgatttttgaagaggttcagagtccttgtaacccttctttaaacc 251
 DB 433 GGGATCATTCACATGATTTTGAAGAGGCTTCAGAGCTCCGTACACCTTTCTTAATCC 492
 OY 252 cttgaagagtggaactacccctcattcaggaacttgatgagacaagctcttccatcag 311
 DB 493 CTTAAGAGAGGAGCACTCTCTATTTCAGAGACTTGAAATGACAAAGCTCTTTTCATCAG 552
 OY 312 acatcagaaggaacttgagcagatttgctcaggaatttaaggaacttgatccatccca 371
 DB 553 ACATCAGAAAGAGACTTGAGCTGCTCAGAGATTAAAGACTTGTACCATACCCCA 612
 OY 372 tcttctcgaactttaccccttgctggaagatgacattatcttcttaactgaaagc 431
 DB 613 TCTTTCTGAACCTTTATCCCTTGCTGGAAGATATGACATTTATTTTAACTTGAAAGC 672
 OY 432 accctcagaagacctgctcctgctggaaggaagaccacaacatccgcgctgagcagctg 491
 DB 673 ACCTTCACAGAACCTGCTGCTGAGGAAGACCAACACATGACCGCTGAGACAGCTG 732
 OY 492 accct 496
 DB 733 ACCCT 737

RESULT 10
 LOCUS CNS01DS8/6
 DEFINITION CNS01DS8 162692 bp DNA PRI 02-MAR-2000
 BAC sequence from the SP64 candidate region at 2p21-2p22 BAC 563N04
 of RPCI-11 library from chromosome 2 of Homo sapiens (Human),
 complete sequence.

ACCESSION AL121658.2 GI:7159619
 VERSION AL121658
 KEYWORDS HTG: HTGS_DRAFT; SP64 genomic DNA interval.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 162692)
 Hazan, J., Fonknechten, N., Mavel, D., Paternotte, C., Sanson, D.,
 Artiguenave, F., Davoine, C.S., Cruaud, C., Durr, A., Wincker, P.,
 Brothier, P., Catolico, L., Barbe, V., Burgeon, J.M.,
 Prud'Homme, J.F., Brice, A., Fontaine, B., Hellig, R. and
 Weissenbach, J.

TITLE Spastin, a novel AAA protein, is altered in the most frequent form
 of autosomal dominant spastic paraplegia
 JOURNAL Nat. Genet. (1999) in press
 REFERENCE 2 (bases 1 to 162692)
 TITLE Direct Submission
 JOURNAL Submitted (29-FEB-2000) to the EMBL/GenBank/DBJ databases
 COMMENT On Mar 6, 2000 this sequence version replaced gi:6002391.
 FEATURES
 source

1. 162692
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /clone_11b="RPCI-11"
 /clone="563N04"
 BASE COUNT 44981 a 32017 c 32887 g 52792 t 15 others
 ORIGIN

Query Match 10.2%: Score 318; DB 9; Length 162692;
 Best Local Similarity 100.0%; Pred. No. 2,6e-160; Indels 0; Gaps 0;
 Matches 318; Conservative 0; Mismatches 0;

OY 2816 aggtgcatcttttgaagagaccttgaaactccacagcattgaattgaggggaa 2875
 DB 152128 AGGTGCATTTTGGAAAGAACCTCTGAAAACTTCACACATTTGAATTTGGCGGAAA 152159
 OY 2876 tgcgtgagcagtgatgagatgctgcctcattcaggtgtatattgagaattcagaact 2935

DB 152158 TCGTGTGACAGCATGATGATGCTGCTTCATGAGGTGTATTGAAATCTTAACCAAT 152099
 OY 2936 aggtgttttgaactttagtactaagaatttctaccctgataccagcatatgacgaact 2995
 DB 152098 AGGTGTGTTTGGACTTTAGTACTAAGAAATTTTACCTGATTCACGATTAAGAAAACT 152039
 OY 2996 taagcaagtgtatcccaagtaactttctcgaagaagcagaacttgcttgaggcaatt 3055
 DB 152038 TACCAAGGTGTATCCAGTTAACTTTTGTGCAAGAAAGCTAGGCTTGTGGTGCCAAAT 151979
 OY 3056 tgatgatgatcctcagtgattatcagaagtgctttaaactagtaactgtcttaataaa 3115
 DB 151978 TGATGATGATGATGATCTCAGTGTATTATACAGTGCTTTTAACTAGTAACTTAAT 151919
 OY 3116 gtgtactcgaagccagta 3133
 DB 151918 GTGTACTCGAAGCCAGTA 151901

RESULT 11
 LOCUS AC011232/6
 DEFINITION AC011232 185281 bp DNA HTG 10-MAR-2001
 Homo sapiens chromosome 2 clone RP11-78E13, WORKING DRAFT SEQUENCE,
 7 unordered pieces.
 ACCESSION AC011232.7 GI:13270720
 VERSION AC011232
 KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 185281)
 Waterston, R.H.
 The sequence of Homo sapiens clone
 unpublished
 2 (bases 1 to 185281)
 Waterston, R.H.
 REFERENCE Direct Submission
 JOURNAL Submitted (04-OCT-1999) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 On Mar 10, 2001 this sequence version replaced gi:9799811.

COMMENT
 ----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: W0GSC
 Web site: http://genome.wustl.edu/gsc/index.shtml
 ----- Project Information -----
 Project name: H.NH0078E13
 Center project name: H.NH0078E13
 ----- Summary Statistics -----

Sequencing vector: M13, 578
 Sequencing vector: plasmid, 408
 Chemistry: Dye-terminator Big Dye; 50% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 182798 bases at least Q40
 Consensus quality: 183538 bases at least Q30
 Consensus quality: 184045 bases at least Q20
 Insert size: 168000; agarose-fp
 Insert size: 184681; sum-of-ctnigs
 Quality coverage: 7.07 in Q20 bases; sum-of-ctnigs
 Quality coverage: 7.33 in Q20 bases; sum-of-ctnigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 7 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N. But the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1
 * 1985: contig of 1985 bp in length
 * 1986
 * 2085: gap of unknown length

*	2086	8183:	contig of 6098 bp in length
*	8184	8283:	gap of unknown length
*	8284	22741:	contig of 14458 bp in length
*	22742	22841:	gap of unknown length
*	22842	41446:	contig of 18605 bp in length
*	41447	41546:	gap of unknown length
*	41547	69182:	contig of 27636 bp in length
*	69183	69282:	gap of unknown length
*	69283	112131:	contig of 42849 bp in length
*	112132	112231:	gap of unknown length
*	112232	185281:	contig of 73050 bp in length

```

/organism:"Homo sapiens"  

/db_xref:"taxon:9606"  

/chromosome:"2"  

/location:"RP11-78E13"  

1..1985  

misc_feature  

/note:"assembly_name:Contig11"  

2086..8183  

misc_feature  

/note:"assembly_name:Contig12"  

8284..122741  

misc_feature  

/note:"assembly_name:Contig13"  

22842..41446  

misc_feature  

/note:"assembly_name:Contig14"  

41547..69182  

misc_feature  

/note:"assembly_name:Contig15"  

69283..112131  

misc_feature  

/note:"assembly_name:Contig16"  

112332..185281  

misc_feature  

/note:"assembly_name:Contig17"  

185281..37187  

BASE COUNT      52719   a    36957   c    37187   g    57816   t    602  

ORIGIN

```

	Query Match	10.2%;	Score 318;	DB 2;	Length 185281;
	Best Local Similarity	100.0%;	Pred. No. 2,6e-160;		
	Matches 318;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0
QY 2816	aggtgcatcttttttggaaagaaacccctctgaaaaactccagcagcttgaatttggcgggaaa				2875
Db 160183	AGGTGCATTTTTTTGGAAAGAACCCCTGTGAAAAACTTCAGCAGTTGAATTTGGCGGGANA				160124
QY 2876	tcgtgtgagcagctgtagtgaatggtcctgtcctcaatgggtgtatttgaagaacctaaagcaatt				2935
Db 160123	TCTGTGTGAGCAGCTGATGTGATGGCTTGCCTTCATGTGGGTGTATTGTGAACCTTAAGCAATT				160064
QY 2936	agtgcttttcttaacttagtaactaaagaatttctaactgttccaagattgttcaagaacct				2995
Db 160063	AGTGTTTTTGTACTTTAGTACTAAAGAAATTTCTACCTGTATCCAGCATTTATGTAGAAAACCT				160004
QY 2996	tagccaagctgtctatccaagtttaacttctctgcaagaagcttagcctgttgggtggcaatt				3055
Db 160003	TAGCCAACTGTTATGCCAAGTTAACTTTTCTGCAAAAGCTTAGCTTTGTGGGTGCCAATT				159944
QY 3056	tgatgatgaatgaactcagctgttattatacagtgctttaaactagtaactgctttaataaa				3115
Db 159943	TGATGATGATGATGATCACTCACTGTTATTATACAGGTCCTTTTAAACTAGTAACCTGTTAATAA				159884
QY 3116	gtgtactactcgaaagccaagta				3133
Db 159883	GTTGACTCTGGAAGCCAGTA				159866

[illegible]

REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 578)
TITLE	Damiano,J.S., Stehlik,C., Pio,F., Godzik,A. and Reed,J.C.
JOURNAL	Clan,a novel human ced-4-like gene
MEDLINE	Genomics. 75 (1-3), 77-83 (2001)
PUBMED	21365712
REFERENCE	11472070
AUTHORS	2 (bases 1 to 578)
TITLE	Stehlik,C., Damiano,J.S., Pio,F., Godzik,A. and Reed,J.C.
JOURNAL	Direct Submission
PUBMED	Submitted (21-Feb-2001)
REFERENCE	Program on Apoptosis and Cell Death

FEATURES	source	location/Qualifiers
		1. .578
		/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/chromosome="2"
		/map="2p22-p21"
		/tissue-type="Lung"
gene		1. .578
		/gene="CLAN1"
		277. .555
CDS		/gene="CLAN1"
		/note="CARD protein"
		/codon_start=1
		/product="CLAND"
		/protein_id="AAK14779.1"
		/db_xref="GI:14324119"
		/translation="MFEIKDQSRRLIORMGTVYKQTTDLFVYNNVLRREVNITCE
		KVEDDANGIITHMLTKGDESCNLEFLSLKEWNPILFDLNGOSLLTA"
BASE COUNT	172 a	106 c 143 g 157 t
ORIGIN		

	Query Match	9.4%	Score 293	DB 9	Length 578
	Best Local Similarity	100.0%	Pred. NO. 1e-146		
	Matches 293	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	12	ggtggaagcttcacacagaacatgatatccataaaggacaatagccagaccctat	71		
Db	253	GGTGGAGGCTTTCATCCAGAACATGATTTTCATTAAGGACAAATAGCCGAGCCTTAT	312		
QY	72	caagaatggaagatgactgttataaagcaaatcacaga tgaacctattgtatgaaagt	131		
Db	313	CAAGAATGGGAATGACTGTTATTAAGCAAAATCACAGATGACCTTTGTATGAGATGTT	372		
QY	132	ctgaatcgagaagaatgaacatcatctgtctgcggaagtgtagcagatgctgtcaga	191		
Db	373	CTGATCCGCGAAGATGAACATCTTTGGCGGAGAGGTGAGACAGATGCTCTAGA	432		
QY	192	gggatcatcatcagatctttgaaaaagggttcagaagtcctgtaacctcttcttaacc	251		
Db	433	GGGATCATTCACATGATTTTGA AAAAGGTTCCAGAGTCTCTGAACCTCTTCTTAAATCC	492		
QY	252	cttaagagtggaatcctcctatcttaagaagctgaaatgagcaaaagctttt	304		
Db	493	CTTAAGGATGGAATATCTCTATTTTGAGGCTTGAAATGGCAAAAGTCTTTT	545		

RESULT	13
LOCUS	G55568/c
DEFINITION	553 bp DNA
ACCESSION	SSHC-100923 Human Homo sapiens STS genomic, sequence tagged site.
VERSION	G55568
KEYWORDS	G55568.1 GI:6120887
SOURCE	STS.
ORGANISM	human.
REFERENCE	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homini; Homidae; Homo. 1 (bases 1 to 553)

AUTHORS Olivier, M. and Cox, D.R.
 TITLE Unpublished, Olivier, M., Cox, D.R. (2000)
 JOURNAL Unpublished (2000)
 COMMENT

Contact: Michael Olivier, David R. Cox
 Stanford Human Genome Center
 Stanford University School of Medicine
 4005 Miranda Ave. 2nd Fl., Palo Alto, CA 94025, USA
 Tel: (650) 320-5800
 Fax: (650) 320-5801
 Email: olivier@shgc.stanford.edu
 Primer A: AATAAGGGGCAAAATAGCAAA
 Primer B: TAACACCCCTTGTCTCTCA
 STS size: 322
 PCR Profile:

Initial incubation: 95 degrees C for 10 minutes
 Denaturation: 94 degrees C for 30 seconds
 Annealing: 60 degrees C for 30 seconds
 Polymerization: 72 degrees C for 23 seconds
 PCR cycles: 30
 Thermal Cycler: Perkin Elmer 9700

Protocol: Template: 25 ng
 Primer: each 1 uM
 dNTPs: each 200 uM
 Amplifrag Gold Polymerase: 0.07 units/ul
 Total Vol: 5 ul

Buffer: MgCl2: 2.5 mM
 KCl: 50 mM
 Tris-HCl: 10 mM
 pH: 8.3

BAC ends sequenced at TIGR from the RPC11 BAC library. Designed and developed at the Stanford Human Genome Center.

FEATURES
 source 1..553
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="2"
 /clone_11b="Human"

STS
 primer_bind 5..27
 primer_bind complement(304..326)
 BASE COUNT 170 a 107 c 114 g 162 t
 ORIGIN

Query Match 5.4%; Score 170; DB 11; Length 553;
 Best Local Similarity 100.0%; Pred. No. 4,6e-80;
 Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 2650 tcgacagatgaacgtgctagaaacagctcaccgactgctgccttggtggtgacg 2709
 |||||||
 Db 258 TCGACGAGTGAACGTGTGAGAACAGCTCACCCGACTGATGCTGCCCTGGGCTGTGAGAC 199
 |||||||
 Y 2710 tgcagaagcgcctgagcagcgtgttgaacatttgagagaggtcccaactcgttaagc 2769
 |||||||
 Db 198 TGCAGAGCAGCCTGAGCAGCCTGTGAAACATTGTGAGAGGAGGTCCACAACTGTCGAAGC 139
 |||||||
 Y 2770 ttgggttgtaaaactgagagactcacagatcacagatattgaattgaatt 2819
 |||||||
 Db 138 TTGGGTTGAAAAAAGCTGAGAGCTCACAGATACAGATTAAGATTATTAGGT 89

RESULT 14
 AC010968 160583 bp DNA HTG 18-AUG-2000
 LOCUS AC010968/c
 DEFINITION Homo sapiens chromosome 2 clone RP11-9302, WORKING DRAFT SEQUENCE,
 11 unoriented pieces.
 AC010968
 AC010968.5 GI:9845170

KEYWORDS
 SOURCE
 ORGANISM

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

HTG; HTGS_PHASE1; HTGS_DRAFT.
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 160583)
 Waterston, R.H.
 The sequence of Homo sapiens clone
 Unpublished
 2 (bases 1 to 160583)
 Waterston, R.H.
 Direct Submission
 Submitted (28-SEP-1999) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 On Aug 18, 2000 this sequence version replaced gi:8439959.

----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc/index.shtml
 ----- Project Information -----
 Center project name: H_NH0093002
 ----- Summary Statistics -----
 Sequencing vector: p13: 59%
 Chemistry: Dye-primer ET; 47% of reads
 Chemistry: Dye-terminator Big Dye; 53% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: bases at least Q40
 Consensus quality: bases at least Q30
 Consensus quality: bases at least Q20
 Insert size: 147000; agarose-fp
 Insert size: 159583; sum-of-contigs
 Quality coverage: 6.64 in Q20 bases; agarose-fp
 Quality coverage: 6.38 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 11 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 1 1080: contig of 1080 bp in length
 * 1081 1180: gap of unknown length
 * 1181 2476: contig of 1296 bp in length
 * 2477 2576: gap of unknown length
 * 2577 5035: contig of 2459 bp in length
 * 5036 5135: gap of unknown length
 * 5136 9606: contig of 4471 bp in length
 * 9607 9706: gap of unknown length
 * 9707 17078: contig of 7372 bp in length
 * 17079 27158: gap of unknown length
 * 27159 27258: contig of 9980 bp in length
 * 27259 27258: gap of unknown length
 * 27259 45137: contig of 17879 bp in length
 * 45138 45237: gap of unknown length
 * 45238 65522: contig of 20285 bp in length
 * 65523 65622: gap of unknown length
 * 65623 91498: contig of 25876 bp in length
 * 91499 91598: gap of unknown length
 * 91599 116835: contig of 25337 bp in length
 * 116836 116935: gap of unknown length
 * 116936 160583: contig of 43648 bp in length.
 Location/Qualifiers
 1..160583
 source
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /clone="RP11-9302"

misc_feature

1..1080

misc_feature /note="assembly_name:Contig9"
1181.2476
/note="assembly_name:Contig12"
misc_feature /note="assembly_name:Contig13"
2577.5035
/note="assembly_name:Contig14"
misc_feature /note="assembly_name:Contig15"
5136.9606
/note="assembly_name:Contig16"
misc_feature /note="assembly_name:Contig17"
9707.17078
/note="assembly_name:Contig18"
misc_feature /note="assembly_name:Contig19"
1719.27158
/note="assembly_name:Contig20"
misc_feature /note="assembly_name:Contig21"
27259.45137
/note="assembly_name:Contig22"
misc_feature /note="assembly_name:Contig23"
45238.65522
/note="assembly_name:Contig24"
misc_feature /note="assembly_name:Contig25"
65623.91498
/note="assembly_name:Contig26"
misc_feature /note="assembly_name:Contig27"
91599.116835
/note="assembly_name:Contig28"
misc_feature /note="assembly_name:Contig29"
116936.160583
/note="assembly_name:Contig30"
misc_feature /note="assembly_name:Contig31"
44349 a 33843 c 33703 g 47641 t 1047 others

BASE COUNT 44349 a 33843 c 33703 g 47641 t 1047 others

ORIGIN

Query Match 0.8% Score 25; DB 2; Length 160583;
Best Local Similarity 100.0%; Pred. NO. 0.15;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 ggtgggaagcttcacagacaa 36
|||||
Db 6875 ggtgggaagcttcacagacaa 6851

RESULT 15
AC074195/c DNA HTG 20-AUG-2000
LOCUS Homo sapiens chromosome 11 clone RP11-750A9, WORKING DRAFT
DEFINITION
AC074195 175152 bp
AC074195.3 GI:9799883
AC074195.3 HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
SOURCE
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 175152)
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS
Waterston, R. H.
JOURNAL
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 175152)
Waterston, R. H.
REFERENCE
Direct Submission
Submitted (16-JUL-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT
On Aug 13, 2000 this sequence version replaced gi:9743495.

Assembly program: Phrap; version 0.990319
Consensus quality: 155181 bases at least Q40
Consensus quality: 161068 bases at least Q30
Consensus quality: 163835 bases at least Q20
Insert size: 174000; agarose-gel
Insert size: 172752; sum-of-contigs
Quality coverage: 4.16 in Q20 bases; agarose-gel
Quality coverage: 4.34 in Q20 bases; sum-of-contigs
NOTE: This is a 'working draft' sequence. It currently
consists of 25 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 1335: contig of 1335 bp in length
1336 1435: gap of unknown length
1436 3074: contig of 1639 bp in length
3075 3174: gap of unknown length
3175 5534: contig of 2360 bp in length
5535 5635: gap of unknown length
5636 8287: contig of 2653 bp in length
8288 8388: gap of unknown length
8389 10415: contig of 2028 bp in length
10416 10515: gap of unknown length
10516 12875: contig of 2360 bp in length
12876 12975: gap of unknown length
12976 16283: contig of 3308 bp in length
16284 16384: gap of unknown length
16385 19547: contig of 3163 bp in length
19548 24020: contig of 4374 bp in length
24021 24120: gap of unknown length
24121 28746: contig of 4626 bp in length
28747 28846: gap of unknown length
28847 33535: contig of 4689 bp in length
33536 33635: gap of unknown length
33636 38255: contig of 4620 bp in length
38256 38355: gap of unknown length
38356 43277: contig of 4922 bp in length
43278 43377: gap of unknown length
43378 47903: contig of 4526 bp in length
47904 52384: gap of unknown length
52385 52484: contig of 4381 bp in length
52485 58975: gap of unknown length
58976 59075: contig of 6491 bp in length
59076 67303: gap of unknown length
67304 67403: contig of 8228 bp in length
67404 74849: gap of unknown length
74850 74949: contig of 7446 bp in length
74950 83615: gap of unknown length
83616 83715: contig of 8666 bp in length
83716 94881: gap of unknown length
94882 94981: contig of 1116 bp in length
94982 104766: gap of unknown length
104767 104866: contig of 9785 bp in length
104867 118097: gap of unknown length
118098 118197: contig of 13231 bp in length
118198 132815: gap of unknown length
132816 132915: contig of 14618 bp in length
132916 148952: gap of unknown length
148953 149052: contig of 16037 bp in length
149053 175152: gap of unknown length
175153: contig of 26100 bp in length.

FEATURES
source
1.175152
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/clone="RP11-750A9"
1.1335

```

misc_feature      /note="assembly_name:Contig13"
                  1436..3074
                  /note="assembly_name:Contig16"
misc_feature      3175..5534
                  /note="assembly_name:Contig17"
misc_feature      5635..8287
                  /note="assembly_name:Contig18"
misc_feature      8388..10415
                  /note="assembly_name:Contig19"
misc_feature      10516..12875
                  /note="assembly_name:Contig20"
misc_feature      12976..16283
                  /note="assembly_name:Contig21"
misc_feature      16384..19546
                  /note="assembly_name:Contig22"
misc_feature      19647..24020
                  /note="assembly_name:Contig23"
misc_feature      24121..28746
                  /note="assembly_name:Contig24"
misc_feature      28847..33535
                  /note="assembly_name:Contig25"
misc_feature      33636..38255
                  /note="assembly_name:Contig26"
misc_feature      38356..43277
                  /note="assembly_name:Contig27"
misc_feature      43378..47903
                  /note="assembly_name:Contig28"
misc_feature      48004..52384
                  /note="assembly_name:Contig29"
misc_feature      52485..58975
                  /note="assembly_name:Contig30"
misc_feature      59076..67303
                  /note="assembly_name:Contig31"
misc_feature      67404..74849
                  /note="assembly_name:Contig32"
misc_feature      74950..83615
                  /note="assembly_name:Contig33"
misc_feature      83716..94881
                  /note="assembly_name:Contig34"
misc_feature      94982..104766
                  /note="assembly_name:Contig35"
misc_feature      104867..118097
                  /note="assembly_name:Contig36"
misc_feature      118198..132815
                  /note="assembly_name:Contig37"
misc_feature      132916..148952
                  /note="assembly_name:Contig38"
misc_feature      149053..175152
                  /note="assembly_name:Contig39"
BASE COUNT      42097 a 43906 c 45616 g 41104 t 2429 others
ORIGIN
Query Match      0.7%; Score 23; DB 2; Length 175152;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1922 agacacagtgatcacatg 1944
Db 80203 AGACACAGGTGATTCACATGG 80181

```

```

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT
-----
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 183556)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 183556)
Waterston,R.H.
Direct Submission
Submitted (30-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Jun 15, 2000 this sequence version replaced gi:7684541.

-----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
-----
Project Information
-----
Center project name: H.NH0125F14
-----
Summary Statistics
-----
Sequencing vector: M13; 70%
Chemistry: Dye-primer ET; 70% of reads
Chemistry: Dye-terminator Big Dye; 30% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 164779 bases at least Q40
Consensus quality: 170994 bases at least Q30
Consensus quality: 174798 bases at least Q20
Insert size: 167000; agarose-fp
Insert size: 180556; sum-of-ctrls
Quality coverage: 3.76 in Q20 bases; agarose-fp
Quality coverage: 3.61 in Q20 bases; sum-of-ctrls
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 31 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1
1181: contig of 1181 bp in length
1182 1281: gap of unknown length
1282 2783: contig of 1502 bp in length
2784 2883: gap of unknown length
2884 4618: gap of 1735 bp in length
4619 4718: gap of unknown length
4719 6318: contig of 1600 bp in length
6319 6418: gap of unknown length
6419 8952: contig of 2534 bp in length
8953 9052: gap of unknown length
9053 11436: contig of 2384 bp in length
11437 11536: gap of unknown length
11537 14108: contig of 2572 bp in length
14109 14208: gap of unknown length
14209 16184: contig of 1976 bp in length
16185 16284: gap of unknown length
16285 20536: contig of 4252 bp in length
20537 20636: gap of unknown length
20637 25264: contig of 4628 bp in length
25265 25364: gap of unknown length
25365 29797: contig of 4433 bp in length
29798 29897: gap of unknown length
29898 33626: contig of 3729 bp in length
33627 33726: gap of unknown length
33727 37441: contig of 3615 bp in length
37442 37441: gap of unknown length
37441 41642: contig of 4201 bp in length
41643 41742: gap of unknown length
41743 47281: contig of 5549 bp in length
47292 47391: gap of unknown length
47392 50826: contig of 3435 bp in length
50827 50926: gap of unknown length

```

```

* 50927 58180: contig of 7254 bp in length
* 58181 58280: gap of unknown length
* 58281 63165: contig of 4885 bp in length
* 63166 63265: gap of unknown length
* 63266 68481: contig of 5216 bp in length
* 68482 73841: gap of unknown length
* 73842 73941: gap of unknown length
* 73942 79919: contig of 5978 bp in length
* 79920 80020: gap of unknown length
* 80020 86821: contig of 6702 bp in length
* 86822 92875: gap of unknown length
* 92876 92977: gap of unknown length
* 92977 98096: contig of 5120 bp in length
* 98097 106663: contig of 8467 bp in length
* 106664 117153: gap of unknown length
* 117154 117254: gap of unknown length
* 117254 128332: contig of 11079 bp in length
* 128333 128433: gap of unknown length
* 128433 139368: contig of 10936 bp in length
* 139369 154140: gap of unknown length
* 154141 154240: contig of 14672 bp in length
* 154241 167723: gap of unknown length
* 167724 167824: contig of 13483 bp in length
* 167824 183556: contig of 15733 bp in length.

```

FEATURES

SOURCE

```

1. 183556
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/clone="RP11-125F14"
1. 1181
/misc_feature /note="assembly_name:Contig7"
1282. 2783
/misc_feature /note="assembly_name:Contig8"
2884. 4618
/misc_feature /note="assembly_name:Contig9"
4719. 6318
/misc_feature /note="assembly_name:Contig10"
6419. 8952
/misc_feature /note="assembly_name:Contig11"
9053. 11436
/misc_feature /note="assembly_name:Contig12"
11537. 14108
/misc_feature /note="assembly_name:Contig13"
14209. 16184
/misc_feature /note="assembly_name:Contig14"
16285. 20536
/misc_feature /note="assembly_name:Contig15"
20637. 25264
/misc_feature /note="assembly_name:Contig16"
25365. 29797
/misc_feature /note="assembly_name:Contig17"
29898. 33626
/misc_feature /note="assembly_name:Contig18"
33727. 37341
/misc_feature /note="assembly_name:Contig19"
37442. 41642
/misc_feature /note="assembly_name:Contig20"
41743. 47291
/misc_feature /note="assembly_name:Contig21"
47392. 50826
/misc_feature /note="assembly_name:Contig22"
50927. 58180
/misc_feature /note="assembly_name:Contig23"
58281. 63165
/misc_feature /note="assembly_name:Contig24"
63266. 68481
/misc_feature /note="assembly_name:Contig25"

```

```

misc_feature 68582. 73841
/note="assembly_name:Contig26"
73942. 79919
/misc_feature /note="assembly_name:Contig27"
80020. 86721
/misc_feature /note="assembly_name:Contig28"
86822. 92876
/misc_feature /note="assembly_name:Contig29"
92977. 98096
/misc_feature /note="assembly_name:Contig30"
98197. 106663
/misc_feature /note="assembly_name:Contig31"
106764. 117153
/misc_feature /note="assembly_name:Contig32"
117254. 128332
/misc_feature /note="assembly_name:Contig33
clone_end:SP6
vector_side:left"
128433. 139368
/misc_feature /note="assembly_name:Contig34
clone_end:T7
vector_side:right"
139469. 154140
/misc_feature /note="assembly_name:Contig35"
154241. 167723
/misc_feature /note="assembly_name:Contig36"
167824. 183556
/misc_feature /note="assembly_name:Contig37"

```

```

BASE COUNT 43865 a 46765 c 45344 g 44562 t 3020 others
ORIGIN

```

```

Query Match 0.7%; Score 23; DB 2; Length 183556;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1922 agacacagtgatccacatgg 1944
|||||
Db 103946 AGACACAGTGGATCCACATCG 103924

```

```

RESULT 17
AC090582 LOCUS
DEFINITION Homo sapiens chromosome 11 clone RP11-125F14 map 11, WORKING DRAFT
ACCESSION AC090582
VERSION AC090582.3 GI:14595831
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 188459)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 11, clone RP11-125F14
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 188459)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguslavsky,L., Bouhagalter,B., Brown,A.,
Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
Collins,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,D.S.,
Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galaan,J.,
Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Harford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Karatas,A., Labocque,K., Lamazares,R., Landers,T.,
Lenczky,J., Levine,R., Liu,G., Maclean,C., Macdonald,P.,
Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
McPheters,R., Meldrum,J., Menes,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
Rieback,M., Riley,K., Rise,C., Rogov,P., Roman,J., Roselli,M.,

```

Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Sounez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Travers, M., Travers, N., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.-J., Young, G., Zalnoun, J., Zemek, L., Zimmer, A. and Zody, M.

TITLE JOURNAL

COMMENT

Submitted (03-MAR-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 4, 2001 this sequence version replaced gl:13603991.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RW/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L11783

Center clone name: 125_F_14

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 182057 bases at least Q40

Consensus quality: 185378 bases at least Q20

Insert size: 192000; agarose-fp

Insert size: 187059; sum-of-ctrls

Quality coverage: 8.4 in Q20 bases; agarose-fp

Quality coverage: 8.6 in Q20 bases; sum-of-ctrls

* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```
1      5671: contig of 5671 bp in length
*      5672 5771: gap of 100 bp
*      5772 6377: contig of 606 bp in length
*      6378 6477: gap of 100 bp
*      6478 7122: contig of 645 bp in length
*      7123 7222: gap of 100 bp
*      7223 8771: contig of 1549 bp in length
*      8772 8871: gap of 100 bp
*      8872 10666: contig of 1795 bp in length
*      10667 10766: gap of 100 bp
*      10767 13839: contig of 3073 bp in length
*      13840 13939: gap of 100 bp
*      13940 18026: contig of 4087 bp in length
*      18027 18126: gap of 100 bp
*      18127 24163: contig of 6037 bp in length
*      24164 24263: gap of 100 bp
*      24264 31828: contig of 7565 bp in length
*      31829 31928: gap of 100 bp
*      31929 69676: contig of 37748 bp in length
*      69677 69776: gap of 100 bp
*      69777 90436: contig of 20660 bp in length
*      90437 90536: gap of 100 bp
*      90537 117488: contig of 26952 bp in length
*      117489 117588: gap of 100 bp
*      117589 145954: contig of 28366 bp in length
*      145955 146054: gap of 100 bp
*      146055 174878: contig of 28824 bp in length
*      174879 174978: gap of 100 bp
*      174979 188459: contig of 13481 bp in length.
```

FEATURES

Source

1. 188459
/organism="Homo sapiens"

```
/db xref="taxon:9606"
/chromosome="11"
/map "11"
/clone="RP11-125F14"
/clone_lib="RPCT-11 Human Male BAC"
1. 5671
/ote="assembly_fragment
clone_end:SP6
vector_side:left"
5772. 6377
/ote="assembly_fragment"
6478. 7122
/ote="assembly_fragment"
7223. 8771
/ote="assembly_fragment"
8872. 10666
/ote="assembly_fragment"
10767. 13839
/ote="assembly_fragment"
13940. 18026
/ote="assembly_fragment"
18127. 24163
/ote="assembly_fragment"
24264. 31828
/ote="assembly_fragment"
31929. 69676
/ote="assembly_fragment"
69777. 90436
/ote="assembly_fragment"
90537. 117488
/ote="assembly_fragment"
117589. 145954
/ote="assembly_fragment"
146055. 174878
/ote="assembly_fragment"
174979. 188459
/ote="assembly_fragment
clone_end:T7
vector_side:right"
BASE COUNT 45628 a 47237 c 48121 g 46068 t 1405 others
ORIGIN
```

```
Query Match 0.7% Score 23; DB 2; Length 188459;
Best Local Similarity 100.0%; Pred No.1.8;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy 1922 agacacagtggaatcacatg 1944

Db 85215 AGACACAGGTGGAATTCACATCG 85237

RESULT 18

AC090559

LOCUS

DEFINITION

SEQUENCE

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

```
AC090559 204487 bp DNA HTG 11-JUL-2001
Homo sapiens chromosome 11 clone RP11-750H9 map 11, WORKING DRAFT
SEQUENCE 16 unordered pieces.
AC090559
AC090559.3 GI:14670098
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 204487)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 11, clone RP11-750H9
Unpublished
2 (bases 1 to 204487)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,
Barna, N., Bastien, V., Boguslavsky, L., Boukhalter, B., Brown, A.,
Camarata, J., Campopiano, A., Chepel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., Deatellano, K., Dewar, K., Diaz, J.S.,
```

TITLE
JOURNAL
COMMENT

Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Haas, B., Hearford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Karatas, A., LaRoque, K., Lamazares, R., Landers, T., Lenoczky, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Marcus, N., Mathews, C., McCarthy, M., McEwan, P., McKernan, K., McNeelers, R., Meldrum, J., Meneus, L., Mihova, T., Mienna, V., Murphy, T., Naylor, J., Nguyen, C., Noddy, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R., Seaman, S., Severy, P., Sougne, C., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Trauers, M., Travis, N., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Submitted (03-MAR-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 11, 2001 this sequence version replaced g1:14150930.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www.seq.wi.mit.edu>Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: 111075

Center clone name: 750.H.9

Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 199468 bases at least Q40
Consensus quality: 201173 bases at least Q30
Consensus quality: 202030 bases at least Q20
Insert size: 200000; agarose-fp
Insert size: 202987; sum-of-contigs
Quality coverage: 10.6 in Q20 bases; sum-of-contigs
Quality coverage: 10.6 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

1      450: contig of 450 bp in length
*      451 550: gap of 100 bp
*      551 3157: contig of 2607 bp in length
*      3158 3257: gap of 100 bp
*      3258 4660: contig of 1403 bp in length
*      4661 4760: gap of 100 bp
*      4761 6875: contig of 2115 bp in length
*      6876 6975: gap of 100 bp
*      6976 9750: contig of 2775 bp in length
*      9751 9850: gap of 100 bp
*      9851 12616: contig of 2766 bp in length
*      12617 12716: gap of 100 bp
*      12717 21088: contig of 8372 bp in length
*      21089 21188: gap of 100 bp
*      21189 27401: contig of 6213 bp in length
*      27402 27501: gap of 100 bp
*      27502 35375: contig of 7874 bp in length
*      35376 35475: gap of 100 bp
*      35476 77415: contig of 41940 bp in length
*      77416 77515: gap of 100 bp
*      77516 91045: contig of 13530 bp in length

```

```

*      91046 91145: gap of 100 bp
*      91146 113825: contig of 22680 bp in length
*      113826 113925: gap of 100 bp
*      113926 137980: contig of 24055 bp in length
*      137981 138080: gap of 100 bp
*      138081 158403: contig of 20323 bp in length
*      158404 158503: gap of 100 bp
*      158504 194539: contig of 36036 bp in length
*      194540 194639: gap of 100 bp
*      194640 204487: contig of 9848 bp in length.

```

FEATURES

source

```

1..204487
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11"
/clone="RP11-750H9"
/clone_lib="RC1-11 Human Male BAC"
1..450
/note="assembly_fragment"
clone_end:sp6
vector_side:left"
551..3157
/note="assembly_fragment"
3258..4660
/note="assembly_fragment"
4761..6875
/note="assembly_fragment"
6976..9750
/note="assembly_fragment"
9851..12616
/note="assembly_fragment"
12717..21088
/note="assembly_fragment"
21189..27401
/note="assembly_fragment"
27502..35375
/note="assembly_fragment"
35476..77415
/note="assembly_fragment"
77516..91045
/note="assembly_fragment"
91146..113825
/note="assembly_fragment"
113926..137980
/note="assembly_fragment"
138081..158403
/note="assembly_fragment"
158504..194539
/note="assembly_fragment"
194640..204487
/note="assembly_fragment"
clone_end:"7"
vector_side:right"

```

```

BASE COUNT      52691 a      48290 c      49218 g      52785 t      1503 others
ORIGIN

```

```

Query Match      0.7%: Score 23; DB 2; Length 204487;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1922 agacacaggtggaatccacatgg 1944
|||||
DB 72357 AGACACAGGTGGAATCCACATGG 72379

```

```

RESULT 19
AP002957 87834 bp. DNA PRI 26-APR-2001
LOCUS Homo sapiens genomic DNA, chromosome 11q, clone:CTD-233717,
DEFINITION complete sequences.
ACCESSION AP002957

```



```

SRS Content:
SHGC-57513 G37313.
Location/Qualifiers
1. 151088
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/clone="CWD-2013NI17"
BASE COUNT 40571 a 38233 c 38536 g 33748 t
ORIGIN

Query Match 0.7% Score 22: DB 9: Length 151088;
Best Local Similarity 100.0%: Pred. No. 6.2;
Matches 22: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1813 attactatctactctcttga 1834
|||||
Db 117449 ATTACTATTGCTCTTTGA 117470

RESULT 21
ID AC007951 standard; DNA; HTG: 151366 BP.
AC AC007951;
SV AC007951.2
DT 04-JUL-1999 (Rel. 60, Created)
DT 05-APR-2000 (Rel. 63, Last updated, Version 3)
DE Homo sapiens chromosome 9 clone RP11-99J1 map 9, WORKING DRAFT SEQUENCE, 9
DE unordered pieces.
XX
XX HTG: HTGS_DRAFT; HTGS_PHASE1.
XX
OS Homo sapiens (human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia
OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.
XX
XX [1]
RP 1-151366
RT "Homo sapiens chromosome 9, clone RP11-99J1";
RL Unpublished.
XX
XX [2]
RP 1-151366
RA Birren B., Linton L., Nusbaum C., Lander E., Allen N., Anderson M.,
RA Baker J., Baldwin J., Barna N., Beckert J., Benn J., Brown A., Castle A.,
RA Cerny J., Collangelo M., Collins S., Collymore A., Cooke P., Dearellano K.,
RA DePavay E., Devon K., Dewar K., Donelan L., Doyle M., Ferreira P.,
RA Fitzhugh W., Forrest C., Funke R., Gage D., Galagan J., Gardyna S.,
RA Gilbert D., Grant G., Hagos B., Heaford A., Horton L., Howland J.C.,
RA Jones C., Kann L., Karatas A., Lehoczy J., Liu C., Locke K.,
RA MacDonald P., Margulis N., McMan P., McGuck A., McKernan K., McLaughlin J.,
RA Meadmond J., Molla M., Morris W., Morrow J., Mychaleckyj J., Naylor J.,
RA Niloff M., O'Connor T., O'Donnell P., Pavlin B., Peterson K., Pollara V.,
RA Riley R., Roberts D., Roy A., Severy P., Strange-Thomann N., Stojanovic N.,
RA Stone K., Sudranthan A., Tesfaye S., Tortuella-Miller I., Vassiliev H.,
RA Vo A., Wagner A., Wheeler J., Wu X., Wyman D., Ye W.J., Zody M.,
RA
RT
RL
RL Submitted (02-JUL-1999) to the EMBL/GenBank/DBJ databases.
RL Whitehead Institute/MIT Center for Genome Research, 320 Charles Street,
RL Cambridge, MA 02141, USA
XX
XX On Apr 1, 2000 this sequence version replaced gi:5332394.
CC All repeats were identified using RepeatMasker:
CC Smit, A.F.A. & Green, P. (1996-1997)
CC http://ftp.genome.washington.edu/RM/RepeatMasker.html
CC
CC ----- Genome Center
CC Center: Whitehead Institute/ MIT Center for Genome Research

```

CC	Center code: WIBR
CC	Web site: http://www.seq.wi.mit.edu
CC	Contact: sequence_submissions@genome.wi.mit.edu
CC	Project Information
CC	Center project name: L878
CC	Center clone name: 99_J_1
CC	Summary Statistics
CC	Sequencing vector: M13; M7815: 96% of reads
CC	Sequencing vector: Plasmid; n/a: 0.0% of reads
CC	3,807/403,245/942/Chemistry: Dye-terminator; Big Dye; 4% of reads
CC	Chemistry: Dye-terminator; Big Dye; 4% of reads
CC	Assembly program: Phrap; version 0.960731
CC	Consensus quality: 13664 bases at least Q40
CC	Consensus quality: 145968 bases at least Q30
CC	Consensus quality: 148942 bases at least Q20
CC	Insert size: 15000; agarose-IP
CC	Insert size: 15056; sum-of-ctf's
CC	Quality coverage:
CC	* NOTE: This is a 'working draft' sequence. It currently
CC	* consists of 9 contigs. The true order of the pieces
CC	* is not known and their order in this sequence record is
CC	* arbitrary. Gaps between the contigs are represented as
CC	* runs of N, but the exact sizes of the gaps are unknown.
CC	* This record will be updated with the finished sequence
CC	* as soon as it is available and the accession number will
CC	* be preserved.
CC	* 1
CC	4870 4969: contig of 4869 bp in length
CC	4870 4969: gap of 100 bp
CC	4970 10428: contig of 5459 bp in length
CC	10429 10528: gap of 100 bp
CC	10529 18565: contig of 8037 bp in length
CC	18566 18665: gap of 100 bp
CC	18666 26259: contig of 7594 bp in length
CC	26260 26359: gap of 100 bp
CC	26360 41183: contig of 14824 bp in length
CC	41184 41283: gap of 100 bp
CC	41284 53957: contig of 12674 bp in length
CC	53958 54057: gap of 100 bp
CC	54058 81728: contig of 27671 bp in length
CC	81729 81828: gap of 100 bp
CC	81829 114786: contig of 32958 bp in length
CC	114787 114886: gap of 100 bp
CC	114887 151366: contig of 36480 bp in length.
XX	Key
XX	Location/Qualifiers
XX	1. 151366
XX	/chromosome="g"
XX	/db_xref="taxon:9606"
XX	/organism="Homo sapiens"
XX	/map="g"
XX	/clone="RP11-99J1"
XX	/clone_lib="RP11-99J1 Human Male BAC"
XX	1. 4869
XX	/note="assembly-fragment-clone_end:SP6-vector_side:right"
XX	4970. 10428
XX	/note="assembly-fragment"
XX	10529. 18565
XX	/note="assembly-fragment"
XX	18666. 26259
XX	/note="assembly-fragment-clone_end:SP6-vector_side:right"
XX	26360. 41183
XX	/note="assembly-fragment"
XX	41284. 53957
XX	/note="assembly-fragment"
XX	54058. 81728
XX	/note="assembly-fragment"
XX	81829. 114786
XX	/note="assembly-fragment"
XX	114887. 151366
XX	/note="assembly-fragment"
XX	Sequence 151366 BP: 46140 A; 28393 C; 28257 G; 47776 T; 800 other;

```

Query Match Similarity      0.7%;   Score 22;   DB 33;   Length 151366;
Best Local Similarity      100.0%;   Pred. No. 6.2;
Matches      22;   Conservative      0;   Mismatches      0;   Indels      0;   Gaps      0

QY      1095      ttccactctcacacacaaacaa 1116
|||||
Db 132511      TTCCACTCTCACACACAAACAA 132532

RESULT  22
AC011980/c
LOCUS
DEFINITION
Homo sapiens clone RP11-16H7, WORKING DRAFT SEQUENCE, 11 unordered
pieces.
ACCESSION
AC011980
VERSION
AC011980.3 GI:7230122
KEYWORDS
HTG; HTGS-PHASE1; HTGS-DRAFT.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 153753)
Biren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens, clone RP11-16H7
Unpublished
2 (bases 1 to 153753)
Biren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Baldwin, J., Barna, N., Becker, R., Boguslavsky, L., Boukhalter, B.,
Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,
Cooke, P., Dekretz, A., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,
Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Lehocky, J., Lien, C., Locke, K., MacDonald, P., Marquis, N.,
McBraw, P., McGuck, A., McKernan, K., McDonald, J., Meidrim, J.,
Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Testa, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (17-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 12, 2000 this sequence version replaced by: 6453966.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

COMMENT
JOURNAL
TITLE
REFERENCE
AUTHORS

* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as

```

```
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
3262: contig of 3262 bp in length
3263: gap of 100 bp
3363: 9412: contig of 6050 bp in length
9413: 9512: gap of 100 bp
9513: 14144: contig of 4632 bp in length
14145: 14244: gap of 100 bp
14245: 24193: contig of 9949 bp in length
24194: 24293: gap of 100 bp
24294: 37092: contig of 12799 bp in length
37093: 37192: gap of 100 bp
37193: 49508: contig of 12316 bp in length
49509: 49608: gap of 100 bp
49609: 60861: contig of 11253 bp in length
60862: 60961: gap of 100 bp
60962: 76206: contig of 15245 bp in length
76207: 76306: gap of 100 bp
76307: 89151: contig of 12845 bp in length
89152: 89251: gap of 100 bp
89252: 107693: contig of 18442 bp in length
107694: 107793: gap of 100 bp
107794: 153733: contig of 45940 bp in length.

FEATURES
    source
        1..153733
            Location/Qualifiers
                organism="Homo sapiens"
                db_xref="taxon:9606"
                clone_lib="RPCI-11 Human Male BAC"
                1..3262
                    /note="assembly-fragment"
                3263..9412
                    /note="assembly-fragment"
                9513..14144
                    /note="assembly-fragment"
                14245..24193
                    /note="assembly-fragment"
                24294..37092
                    /note="assembly-fragment"
                37193..49508
                    /note="assembly-fragment"
                49609..60861
                    /note="assembly-fragment"
                60909..60861
                    /note="assembly-fragment"
                clone_end:T7
                vector_side:right"
                60962..76206
                    /note="assembly-fragment"
                76307..89151
                    /note="assembly-fragment"
                clone_end:SP6
                vector_side:right"
                89252..107693
                    /note="assembly-fragment"
                107794..153733
                    /note="assembly-fragment"

BASE COUNT      42218 a 34152 c 33610 g 42710 t      1043 others
ORIGIN
```

```
Query Match      0.7%: Score 22; DB 2; Length 153733;
Best Local Similarity 100.0%: Pred. No. 6.2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Oy 1152 aaaaacaacaacaataag 1173
|||||

Db 120518 AAAAACAAACAAACATAAG 120497

RESULT 23
AL592213/c
LOCUS AL592213 155531 bp DNA HTG 04-AUG-2001

```
DEFINITION Homo sapiens chromosome 9 clone RP11-99J1, *** SEQUENCING IN
PROGRESS ***, in ordered pieces.
ACCESSION AL592213 AC007951
VERSION AL592213.6 GI:15131900
KEYWORDS HTG; HTGS_PHASE2; HTGS_CANCELLED.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 155531)
AUTHORS Hammond, S.
TITLE Direct Submissions
JOURNAL Submitted (03-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
COMMENT On Aug 9, 2001 this sequence version replaced gi:15029558.
Draft Sequence produced by Whitehead Institute/MIT Center for
Genome Research, 320 Charles Street,
Cambridge, MA 02141, USA
http://www-seq.wi.mit.edu

----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: BA99J1
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: M13; M77815; 30% of reads
Sequencing vector: plasmid; L08752; 69% of reads
Chemistry: Dye-terminator ET-amersham; 3% of reads
Chemistry: Dye-terminator Big Dye; 68% of reads
Chemistry: Dye-terminator-amersham; 28% of reads
Consensus quality: 155436 bases at least Q40
Consensus quality: 155453 bases at least Q30
Consensus quality: 155482 bases at least Q20
Insert size: 155531; sum-of-contigs
Insert size: 144681; 16.2% error; agarose-fp
Quality coverage: 15.49x in Q20 bases; sum-of-contigs Quality
coverage: 16.65x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
    source
        1..155531
            Location/Qualifiers
                organism="Homo sapiens"
                db_xref="taxon:9606"
                chromosome="9"
                clone_lib="RPCI-11.1"
                1..155531
                    /note="assembly-fragment:03518"
                clone_end:SP6
                vector_side:right"

BASE COUNT      50143 a 29768 c 28721 g 46899 t
ORIGIN
```

```
Query Match      0.7%: Score 22; DB 2; Length 155531;
Best Local Similarity 100.0%: Pred. No. 6.2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Oy 1095 ttccactctcacacacaaca 1116
|||||

Db 38569 TTCACCTCTCACACACAA 38548

RESULT 24
AP002958/c
LOCUS AP002958 156230 bp DNA HTG 05-DEC-2000

Db 73708 AAAACAAACACACATTAAG 73687

|||||

RESULT 25
AC079194/C
LOCUS
DEFINITION
AC079194 159946 bp DNA HTG 22-NOV-2000
SEQUENCE, 37 unordered pieces.
AC079194
AC079194.2 GI:11276162
HTG: HTGS_PHASE1, HTGS_DRAFT.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
2 (bases 1 to 159946)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Beda, F., Boguslavsky, L.,
Bouckgalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A.,
Choepl, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
DeRellano, K., Dewar, K., Diaz, J.S., Dodge, S., Ferreira, P.,
Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M.,
Graham, L., Grand-Pierre, N., Hagos, B., Heatford, A., Horton, L.,
Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., LaCocque, K.,
Lamavez, R., Landers, T., Lechoczy, J., Levine, R., Lieu, C., Liu, G.,
Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McKernan, K.,
McPheeters, R., Meldrum, J., Meneus, L., Mihova, T., Mlenga, Y.,
Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T.,
O'Donnell, P., O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K.,
Pierre, N., Pisanic, C., Pollara, V., Raymond, C., Rieback, M., Riley, R.,
Rogov, P., Roehman, D., Roy, A., Santos, R., Schauer, S., Severy, P.,
Sougnier, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A.,
Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J.,
Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT
Direct Submission
Submitted (23-AUG-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Nov 22, 2000 this sequence version replaced gi:9886030.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIRR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: I10601

Center clone name: 475.C.4

Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 139893 bases at least Q40

Consensus quality: 150490 bases at least Q30

Consensus quality: 154263 bases at least Q20

Insert size: 17600; agarose-1p

Insert size: 156346; sum-of-contigs

Quality coverage: 3.2 in Q20 bases; agarose-1p

Quality coverage: 3.6 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 37 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
2598 2697: gap of 100 bp
2698 3739: contig of 1042 bp in length
3740 3839: gap of 100 bp
3840 4797: contig of 958 bp in length
4798 4897: gap of 100 bp
4898 6197: contig of 1300 bp in length
6198 6297: gap of 100 bp
6298 8221: contig of 1924 bp in length
8222 8321: gap of 100 bp
8322 9558: contig of 1257 bp in length
9559 9658: gap of 100 bp
9659 11181: contig of 1523 bp in length
11182 11281: gap of 100 bp
11282 12622: contig of 1341 bp in length
12623 12722: gap of 100 bp
12723 14374: contig of 1652 bp in length
14375 14474: gap of 100 bp
14475 37771: contig of 23297 bp in length
37772 37871: gap of 100 bp
37872 39354: contig of 1483 bp in length
39355 39454: gap of 100 bp
39455 40854: contig of 1400 bp in length
40855 40954: gap of 100 bp
40955 42709: contig of 1755 bp in length
42710 42809: gap of 100 bp
42810 44167: contig of 1358 bp in length
44168 44267: gap of 100 bp
44268 46404: contig of 2137 bp in length
46405 46504: gap of 100 bp
46505 48310: contig of 1806 bp in length
48311 48410: gap of 100 bp
48411 51136: contig of 2726 bp in length
51137 51236: gap of 100 bp
51237 53831: contig of 2595 bp in length
53832 53931: gap of 100 bp
53932 56439: contig of 2508 bp in length
56440 56539: gap of 100 bp
56540 58998: contig of 2459 bp in length
58999 59098: gap of 100 bp
59099 63491: contig of 4333 bp in length
63492 63591: gap of 100 bp
63592 66743: contig of 3152 bp in length
66744 66843: gap of 100 bp
66844 69817: contig of 2974 bp in length
69818 69917: gap of 100 bp
69918 73103: contig of 3186 bp in length
73104 73203: gap of 100 bp
73204 77583: contig of 4380 bp in length
77584 77683: gap of 100 bp
77684 83310: contig of 5627 bp in length
83311 83410: gap of 100 bp
83411 87029: contig of 3619 bp in length
87030 87129: gap of 100 bp
87130 92085: contig of 4956 bp in length
92086 92185: gap of 100 bp
92186 97647: contig of 5462 bp in length
97648 97747: gap of 100 bp
97748 101467: contig of 3720 bp in length
101468 101567: gap of 100 bp
101568 106778: contig of 5211 bp in length
106779 106878: gap of 100 bp
106879 114040: contig of 7152 bp in length
114041 114140: gap of 100 bp
114141 123890: contig of 9750 bp in length
123891 123990: gap of 100 bp
123991 130670: contig of 6680 bp in length
130671 130770: gap of 100 bp
130771 139820: contig of 9050 bp in length
139821 139920: gap of 100 bp

```
*      *      *
158921 158959: contig of 19039 bp in length
158960 159059: gap of 100 bp
159060 159946: contig of 887 bp in length.
FEATURES
source
1. 159946
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11"
/clone="RP11-475C4"
/clone_lib="RPC1-11 Human Male BAC"
misc_feature
1. 2597
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
2698. 3739
/note="assembly_fragment"
misc_feature
3840. 4797
/note="assembly_fragment"
misc_feature
4898. 6197
/note="assembly_fragment"
misc_feature
6298. 8221
/note="assembly_fragment"
misc_feature
8322. 9558
/note="assembly_fragment"
misc_feature
9659. 11181
/note="assembly_fragment"
misc_feature
11282. 12622
/note="assembly_fragment"
misc_feature
12723. 14374
/note="assembly_fragment"
misc_feature
14475. 37771
/note="assembly_fragment"
misc_feature
37872. 39354
/note="assembly_fragment"
misc_feature
39455. 40854
/note="assembly_fragment"
misc_feature
40955. 42709
/note="assembly_fragment"
misc_feature
42810. 44167
/note="assembly_fragment"
misc_feature
44268. 46404
/note="assembly_fragment"
misc_feature
46505. 48310
/note="assembly_fragment"
misc_feature
48411. 51136
/note="assembly_fragment"
misc_feature
51237. 53831
/note="assembly_fragment"
misc_feature
53932. 56439
/note="assembly_fragment"
misc_feature
56540. 58998
/note="assembly_fragment"
misc_feature
59099. 63491
/note="assembly_fragment"
misc_feature
63592. 66743
/note="assembly_fragment"
misc_feature
66844. 69817
/note="assembly_fragment"
misc_feature
69918. 73103
/note="assembly_fragment"
Query Match
Best Local Similarity 100.0%; Score 22; DB 2; Length 159946;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
RESULT 26
AL161913/CG AL161913 186413 bp DNA HTG 11-JUL-2001
```

```
OY 1152 aaaaacaacaacaataag 1173
|||||
Db 150590 AAAACAACAACAATAAG 150569
```

```
DEFINITION
Homo sapiens chromosome 9 clone RP11-64P11, *** SEQUENCING IN
PROGRESS ***, 2 unordered pieces.
ACCESSION
AL161913
VERSION
AL161913.10 GI:14329900
KEYWORDS
HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 186413)
REFERENCE
Mashreghi-Mohammadi, M.
Direct Submission
Submitted (10-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jun 8, 2001 this sequence version replaced gi:14141268.
COMMENT
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: DB64P11
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator ET-amersham; 7% of reads
Dye-terminator Big Dye; 92% of reads
Consensus quality: 185600 bases at least Q40
Consensus quality: 185986 bases at least Q30
Consensus quality: 186182 bases at least Q20
Insert size: 140000; 37.9% error; agarose-fp
Quality coverage: 9.01x in Q20 bases; sum-of-contigs quality
coverage: 12.46x in Q20 bases; agarose-fp
```

```
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
```

```
* 1 168456: contig of 168456 bp in length
* 168457 168556: gap of 100 bp
* 168557 186413: contig of 17857 bp in length.
```

FEATURES

source

```
1. 186413
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-64P11"
/clone_lib="RPC1-11.1"
1. 168456
/note="assembly_fragment:00638
clone_end:SP6
vector_side:left"
misc_feature
168557. 186413
/note="assembly_fragment:00456"
```

```
BASE COUNT 60691 a 35149 c 33874 g 56599 t 100 others
ORIGIN
```

```
Query Match
Best Local Similarity 100.0%; Score 22; DB 2; Length 186413;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1095 ttccactctcacacaacaacaa 1116
|||||
Db 174706 TTCACCTCTCACACCAACAA 174685
```

RESULT 27
AL592438/c
LOCUS
DEFINITION Homo sapiens chromosome 9 clone RP11-187G6, *** SEQUENCING IN
PROGRESS ***, in ordered pieces.
ACCESSION AL592438
VERSION AL592438.6 GI:15021052
KEYWORDS HTG: HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
AUTHORS Wall,M.
TITLE Direct Submission
JOURNAL Submitted (01-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jul 25, 2001 this sequence version replaced gi:14787678.
COMMENT
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: ba187G6
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: M13; M7815; 9% of reads
Sequencing vector: plasmid; L08752; 90% of reads
Chemistry: Dye-terminator Big Dye; 90% of reads
Chemistry: Dye-primer-amersham; 9% of reads
Consensus quality: 194192 bases at least Q40
Consensus quality: 194424 bases at least Q30
Consensus quality: 194498 bases at least Q20
Insert size: 194533; sum-of-ctrligs
Insert size: 166314; agarose-fp
Quality coverage: 9.0ix in Q20 bases; sum-of-ctrligs Quality
coverage: 10.54x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
1. 194533
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-187G6"
/clone.lib="RPCT-11.1"
misc_feature
1. 194533
/note="assembly_fragment:02981"
SE COUNT 61649 a 36707 c 36581 g 59596 t
IGIN
Query Match 0.7%; Score 22; DB 2; Length 194533;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1095 ttccactctcacacacacaa 1116
|||||
Db 194370 TTCACACTCTCACACACAA 194349
RESULT 28
BTCSK35
LOCUS
DEFINITION BRCASK35 7595 bp DNA
ACCESSION X14908 X14326
VERSION X14908.1 GI:180
KEYWORDS casein; kappa-casein; repetitive sequence: Alu-like repetitive

SOURCE
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE
AUTHORS Alexander,L.J., Stewart,A.F., Mackinlay,A.G., Kapelinskaya,T.V.,
Trach,T.M. and Gorodetsky,S.I.
TITLE Isolation and characterization of the bovine kappa-casein gene
JOURNAL Eur. J. Biochem. 178 (2), 395-401 (1988)
MEDLINE 89091174
REFERENCE
AUTHORS Vaiman,D., Mercier,D., Moazami-Goudarzi,K., Eggen,A.,
Ciampolini,R., Leplinge,A., Velnala,R., Kukinen,J., Varvio,S.L.,
Martin,P. et.al.
TITLE A set of 99 cattle microsatellites: characterization, synteny
mapping, and polymorphism
JOURNAL Mamm. Genome 5 (5), 288-297 (1994)
MEDLINE 94355772
REFERENCE
AUTHORS Mackinlay,A.G.
TITLE Direct Submission
JOURNAL Submitted (16-MAR-1989) A.G. Mackinlay, The University of New South
Wales, P.O. Box 1 Kensington New South Wales Australia
location/Qualifiers
1. 7595
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone.lib="EMBL3, EMBL4."
2886. 2891
/label=stop
/usedin=X14907:kc_s1g
2886. 2918
/label=ex3
/usedin=X14907:kc_cds
/usedin=X14906:kc_mrna
join(2892.. 2918,4930.. 5412)
/label=kc.mat
/product="kappa-casein"
4930. 5412
/label=stop
/usedin=X14907:kc_cds
4930. 5446
/label=ex4
/usedin=X14906:kc_mrna
5309
/note="c is t in kappa-casein B variant; changes acc (Thr)
to atc (Ile)"
5345
/note="a is c in kappa-casein B variant; changes gat (Asp)
to gct (Ala)"
5406
/note="a is g in kappa-casein B variant; loss of pslr
site"
5413
/note="a is t in kappa-casein B variant"
7296. 7416
/label=ex5
/usedin=X14906:kc_mrna
7382
/note="t is c in kappa-casein B variant; loss of BglII
site, creates MspI site"
7416
7464
/note="t is c in variant clone"
7486
/note="t is c in variant clone"
7504
/note="t is c in variant clone"
7554
/note="g is a in variant clone"
7570. 7571
variation
polysite
variation
variation
variation
variation
variation
variation

/note="c is inserted in variant clone"
BASE COUNT 2646 a 1292 c 1182 g 2475 t
ORIGIN

Query Match 0.7%; Score 21; DB 4; Length 7595;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 415 ttctaactgaaagcaccct 435
|||||
DB 6007 TTTTAACTGAAAGCACCCT 6027

RESULT 29
LOCUS AC005199
DEFINITION Homo sapiens chromosome 16, cosmid clone 330D11 (LANL), complete sequence.
ACCESSION AC005199
VERSION AC005199.1 GI:3273387
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 37906)
Ricke,D.O., Bruce,D., Muntz,M., Doggett,N., Munk,C., Saunders,E., Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L., Bryant,J., Tesmer,J., Melnick,L., Longmire,J., White,S., Ueng,S., Tatum,O., Campbell,C., Fawcett,J., Maltbie,M., Misra,M. and Deaven,L.
TITLE Sequencing of Human Chromosome 16p13.3
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 37906)
Ricke,D.O.
TITLE Large Scale Sequence Analysis and Annotation with the Sequence Comparison Analysis (SCAN) System
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 37906)
Ricke,D.O., Bruce,D., Muntz,M., Doggett,N., Munk,C., Saunders,E., Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L., Bryant,J., Tesmer,J., Melnick,L., Longmire,J., White,S., Ueng,S., Tatum,O., Campbell,C., Fawcett,J., Maltbie,M., Misra,M. and Deaven,L.
TITLE Direct Submission
JOURNAL Submitted (30-JUN-1998) Center for Human Genome Studies, DOE Joint Genome Institute, Los Alamos National Laboratory, MS M888, LOS Alamos, NM 87545, USA
FEATURES
source
1. 37906
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="330D11"
/chromosome="16"
/map="16p13.3"
2548..2792
/rpt_family="MLT1"
4130..4301
/rpt_family="Alu"
5671..5997
/note="97% identity EST yb12f02.r1"
/db_xref="dbEST:T47043"
repeat_region
6373..6501
/rpt_family="MER25"
6572..6858
/rpt_family="Alu"
7074..7299
/rpt_family="L1"
7509..7828
/rpt_family="L1"
7898..10788
/rpt_family="L1"
repeat_region

repeat_region 11765..11854
/rpt_family="MER21"
repeat_region complement(12556..12682)
/rpt_family="MER2"
repeat_region complement(13150..13299)
/rpt_family="L1"
repeat_region 13754..14096
/rpt_family="Alu"
repeat_region 14119..14293
/rpt_family="MER5"
repeat_region complement(14450..14651)
/rpt_family="MER"
repeat_region 15660..15985
/rpt_family="Alu"
repeat_region complement(17523..17908)
/rpt_family="MER2a"
repeat_region 18375..18547
/rpt_family="MER"
misc_feature complement(18614..18884)
/note="GRAIL 2 excellent exon, frame 0"
repeat_region 19269..19921
/rpt_family="MER44C"
repeat_region 20666..27120
/rpt_family="L1"
repeat_region 22343..22577
/rpt_family="MER25"
repeat_region 30116..30259
/rpt_family="HSAT1"
repeat_region complement(30385..30580)
/rpt_family="HSAT1"
repeat_region complement(30678..30786)
/rpt_family="HSAT1"
repeat_region complement(31921..32340)
/rpt_family="MLT1"
repeat_region 33481..33615
/rpt_family="Alu"
repeat_region 33613..34213
/rpt_family="L1"
repeat_region 34664..34710
/rpt_family="MIR"
misc_feature complement(34811..35200)
/note="99% identity A0009624 BAC end sequence"
35732..36036
/rpt_family="Alu"
36016..36036
/note="(A)21"
/rpt_type=tandem
/rpt_unit=A
repeat_region complement(36219..36374)
/rpt_family="MLT1"
BASE COUNT 13076 a 7620 c 7679 g 9531 t
ORIGIN

Query Match 0.7%; Score 21; DB 9; Length 37906;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2012 caggactctgagagtcaccct 2032
|||||
DB 1512 CAGGACTCTGAGAGTCACACT 1532

RESULT 30
LOCUS AC017948
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION AC017948
VERSION AC017948.1 GI:6553242
KEYWORDS HTG; HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster

REFERENCE
TITLE
AUTHORS
JOURNAL

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 43496)
Adams, M. and Venter, J.C.
Direct Submission
Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA

COMMENT

This sequence was identified as CDK:10212758 by the submitter. For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES

source
1..45496
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"

BASE COUNT 13595 a 9140 c 9064 g 13697 t
ORIGIN

Query Match 0.7%; Score 21; DB 2; Length 45496;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1666 aaattctgaagccataaaca 1686
29997 AAATCTGAAGCCATTAACA 29977

RESULT 31
LOCUS AC027055/c
DEFINITION Homo sapiens chromosome 3 clone RP11-334K8 map 3, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC027055
VERSION AC027055.1 GI:7329416
KEYWORDS HTG: HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 76526)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 3, clone RP11-334K8
Unpublished
2 (bases 1 to 76526)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F., Boguski, L., Bougaltier, B., Brown, A., Burkett, G., Campilano, A., Castle, A., Choepel, I., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearlano, K., Dewar, K., Diaz, J.S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hages, B., Heathford, A., Horton, L., Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Lacroque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lien, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McGuire, A., McKernan, K., McPheters, R., Melidim, J., Menes, L., Mihova, T., Miranda, C., Mienna, V., Morrow, J., Murphy, T., Naylor, T., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Norman, C.H., O'Connor, T., O'Donnell, P., Pisanil, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Kochman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange, Thoman, N., Stojanovic, N., Subramanian, A., Talmas, J., Testave, S., Theodore, J., Tirrell, A., Travers, M., Triggilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT

Direct Submission
Submitted (26-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence.submissions@genome.wi.mit.edu

Project Information
Center project name: L7376
Center clone name: 334_K_8

* NOTE: This record contains 86 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 803: contig of 803 bp in length
804 903: gap of 100 bp
904 1690: contig of 787 bp in length
1691 1790: gap of 100 bp
1791 2559: contig of 769 bp in length
2560 2659: gap of 100 bp
2660 3445: contig of 786 bp in length
3446 3545: gap of 100 bp
3546 4348: contig of 803 bp in length
4349 4448: gap of 100 bp
4449 5243: contig of 795 bp in length
5244 5343: gap of 100 bp
5344 6119: contig of 776 bp in length
6120 6219: gap of 100 bp
6220 6998: contig of 779 bp in length
6999 7098: gap of 100 bp
7099 7884: contig of 786 bp in length
7885 7984: gap of 100 bp
7985 8766: contig of 782 bp in length
8767 8866: gap of 100 bp
8867 9669: contig of 803 bp in length
9670 9769: gap of 100 bp
9770 10559: contig of 790 bp in length
10560 10659: gap of 100 bp
10660 11449: contig of 790 bp in length
11450 11549: gap of 100 bp
11550 12330: contig of 781 bp in length
12331 12430: gap of 100 bp
12431 13222: contig of 792 bp in length
13223 13322: gap of 100 bp
13323 14098: contig of 776 bp in length
14099 14198: gap of 100 bp
14199 14992: contig of 794 bp in length
14993 15092: gap of 100 bp
15093 15881: contig of 789 bp in length
15882 15981: gap of 100 bp
15982 16771: contig of 790 bp in length
16772 16871: gap of 100 bp
16872 17655: contig of 784 bp in length
17656 17755: gap of 100 bp
17756 18549: contig of 794 bp in length
18550 18649: gap of 100 bp
18650 19440: contig of 791 bp in length
19441 19540: gap of 100 bp
19541 20335: contig of 795 bp in length
20336 20435: gap of 100 bp
20436 21236: contig of 801 bp in length
21237 21336: gap of 100 bp
21337 22140: contig of 804 bp in length
22141 22240: gap of 100 bp
22241 23041: contig of 801 bp in length

```

* 23042 23141: gap of 100 bp
* 23142 23942: contig of 801 bp in length
* 23943 24042: gap of 100 bp
* 24043 24040: contig of 798 bp in length
* 24841 24940: gap of 100 bp
* 24941 25729: contig of 789 bp in length
* 25730 25829: gap of 100 bp
* 25830 26603: contig of 774 bp in length
* 26604 26703: gap of 100 bp
* 26704 27480: contig of 777 bp in length
* 27481 27580: gap of 100 bp
* 27581 28381: contig of 801 bp in length
* 28382 28481: gap of 100 bp
* 28482 29279: contig of 798 bp in length
* 29280 29379: gap of 100 bp
* 29380 30167: contig of 788 bp in length
* 30168 30267: gap of 100 bp
* 30268 31056: contig of 789 bp in length
* 31057 31156: gap of 100 bp
* 31157 31955: contig of 799 bp in length
* 31956 32055: gap of 100 bp
* 32056 32844: contig of 789 bp in length
* 32845 32944: gap of 100 bp
* 32945 33731: contig of 787 bp in length
* 33732 33831: gap of 100 bp
* 33832 34582: contig of 751 bp in length
* 34583 34682: gap of 100 bp
* 34683 35480: contig of 798 bp in length
* 35481 35580: gap of 100 bp
* 35581 36381: contig of 801 bp in length
* 36382 36481: gap of 100 bp
* 36482 37276: contig of 795 bp in length
* 37277 37376: gap of 100 bp
* 37377 38182: contig of 806 bp in length
* 38183 38282: gap of 100 bp
* 38283 39072: contig of 790 bp in length
* 39073 39172: gap of 100 bp
* 39173 39983: contig of 811 bp in length
* 39984 40083: gap of 100 bp
* 40084 40873: contig of 780 bp in length
* 40874 40973: gap of 100 bp
* 40974 41759: contig of 786 bp in length
* 41760 41859: gap of 100 bp
* 41860 42652: contig of 793 bp in length
* 42653 42752: gap of 100 bp
* 42753 43560: contig of 808 bp in length
* 43561 43660: gap of 100 bp
* 43661 44448: contig of 788 bp in length
* 44449 44548: gap of 100 bp
* 44549 45335: contig of 787 bp in length
* 45336 45435: gap of 100 bp
* 45436 46220: contig of 785 bp in length
* 46221 46320: gap of 100 bp
* 46321 47108: contig of 788 bp in length
* 47109 47208: gap of 100 bp
* 47209 47983: contig of 775 bp in length
* 47984 48083: gap of 100 bp
* 48084 48863: contig of 780 bp in length
* 48864 48963: gap of 100 bp
* 48964 49755: contig of 792 bp in length
* 49756 49855: gap of 100 bp
* 49856 50659: contig of 804 bp in length
* 50660 50759: gap of 100 bp
* 50760 51542: contig of 783 bp in length
* 51543 51642: gap of 100 bp
* 51643 52427: contig of 785 bp in length
* 52428 52527: gap of 100 bp
* 52528 53328: contig of 801 bp in length
* 53329 53428: gap of 100 bp
* 53429 54211: contig of 783 bp in length
* 54212 54311: gap of 100 bp
* 54312 55088: contig of 777 bp in length
* 55089 55188: gap of 100 bp

```

```

Query Match      0.7%; Score 21; DB 2; Length 76526;
Best Local Similarity 100.0%; Pred.No. 22;
Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1146 atacagaaacacacacacaa 1166
Db 4183 ATACGAAAAACACACAAA 4163
|||||

```

```

RESULT 32
ATM4E13
LOCUS
DEFINITION Arabidopsis thaliana DNA chromosome 4, BAC clone M4E13 (ESSAII
AL022023
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

```

```

thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 80346)
Bevan,M., Purnelle,B., Boutry,M., Goffeau,A., Hohelsel,J.,
Mewes,H.W., Mayer,K.F.X. and Schueller,C.
Unpublished
2 (bases 1 to 80346)
EU Arabidopsis sequencing/project.

```

```

REFERENCE
AUTHORS
TITLE
JOURNAL
Submitted (04-APR-1998) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, Project
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.bevan@bbsrc.ac.uk

```

FEATURES

```

source
1..80346
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
complement(join(2..154,234..572))
/gene="M4E13.10"
2..572
complement(join(<2..154,234..572))
/gene="M4E13.10"
/gene="M4E13.10"
complement(join(<2..154,234..572))
/notice="similarity to Arabidopsis thaliana AT.I.24-7,
PATCH:G1532169"
/codon_start=1
/product="EF-1 alpha-like protein (fragment)"
/protein_id="CA17760.1"
/db_xref="GI:445155"
/db_xref="SPTREMBL:049604"
/translation="MGFGTSSSSSSSALKMLGFTAVWVQSSGNNYTFSSGALK
SLNMLTLELNSLSVARDVGRKIFILGLASDRSTFVILIGSFEGILGYVOMLVY
SRTPQPIPYQMVCVFLCMGNSMTWMNAVLYVTCIRNRRNRGVSILKGYVLS7A
IFND"
complement(2..154)
exon

```


gene
/gene="M4E13.50"
/number=2
complement(join(20820..21014,21397..21499,21644..21723,
21807..21889,22003..22051,22140..22205,22752..22801,
22966..23029,23242..23354))
/gene="M4E13.60"
complement(join(20820..21014,21397..21499,21644..21723,
22966..23029,23242..23354))

CDS
Query Match 0.7%; Score 21; DB 8; Length 80346;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 626 caaatgctctctctctccg 646
|||||

db 75913 CAAATTCGCTCTCTCCG 75933

RESULT 33

LOCUS ATT12J5 84499 bp DNA PLN 24-FEB-1999
DEFINITION Arabidopsis thaliana DNA chromosome 4, BAC clone T12J5 (ESSAII
project).
ACCESSION AL035522
VERSION AL035522.1 GI:4455339
KEYWORDS
SOURCE
ORGANISM
Chale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 84499)
Beyan,M., Murphy,G., Ridley,P., Hudson,S., Hobeisel,J., Mewes,H.W.,
Mayer,K.F.X. and Schueller,C.
Unpublished
2 (bases 1 to 84499)
EU Arabidopsis sequencing project.
Direct Submission
Submitted (22-FEB-1999) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
schueller@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
Coordinator: Mike Beyan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.beyan@jbsrc.ac.uk
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosome 4 can be
viewed at: <http://webserv.mips.biochem.mpg.de/proj/thal/>.

FEATURES
source
1..84499
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
1..49380
/note="position 1-49380 overlaps to BAC clone M4E13, EMBL
acc:AL022023; for sequence analysis and annotation please
refer to this entry"

misc_feature
1..49380
/note="position 1-49380 overlaps to BAC clone M4E13, EMBL
acc:AL022023; for sequence analysis and annotation please
refer to this entry"

gene
48992..51335
/gene="T12J5.10"
complement(join(48992..49360,49762..50493,50815..51066,
51207..51335))
/gene="T12J5.10"
complement(join(48992..49360,49762..50493,50815..51066,
51207..51335))
/note="similarity to Human (H326) mRNA, Homo sapiens,
gb:U06631"
/codon_start=1
/product="putative protein"
/protein_id="CAB36721.1"
/db_xref="GI:4455340"
translation="MSKARRGRSGNGLHHVVDVDFMRREVGGISRNFSNRFSASEN
LVRLLEYKLEKKKGVNMYSPNAEDGVLLSGDRRVYLMDQMLGKATSPISGHA
NNYQAKFMPSDDRTIYTCAGDMFDLRTAPTELTFCRSVDPRRRMKAIOLEAIA

IDPNSNLFAVGMEVARYLDIRRFOEGINGFTRAADHFCPPHLIGNEDVGTGLA
FSQSEELVSNDEFILFTPGMIGSPISPIKSPVSKSESSSPKDEHSVS
LVYKGRKCEYKGVNFGFSGVSGDGRITMKKSGELIRMEADRHVNCI
EPHPHIVPIASSGIESDIKVMYSKAAEATPELIELSPRPIPMVSLFRRHVDDE
LFGNMDGIGIGNEBDESIDDDAEDNDDSDYSVSGVLLDDNSDDDDDDDDDD
DSDDCEIETIDNNNNNGDGGSETNVSISGSHDDDDDDDDDD
complement(48992..49360)
/gene="T12J5.10"
/number=1
complement(49361..49761)
/number=1
complement(49762..50493)
/gene="T12J5.10"
/number=2
complement(50494..50814)
/number=2
complement(50815..51066)
/gene="T12J5.10"
/number=3
complement(51067..51206)
/number=3
complement(51207..51335)
/gene="T12J5.10"
/number=4
complement(join(51806..52129,52611..52939,53065..53389))
/gene="T12J5.20"
complement(51806..52129,52611..52939,53065..53389))
/gene="T12J5.20"
complement(51806..52129,52611..52939,53065..53389))
/number=1
51806..53389
/gene="T12J5.20"
complement(join(51806..52129,52611..52939,53065..53389))
/note="strong similarity to caffeic acid
O-methyltransferase, Pinus taeda, gb:U39301"
/codon_start=1
/product="O-methyltransferase-like protein"
/protein_id="CAB36722.1"
/db_xref="GI:4455341"
translation="MEESKRNILDEEAKASLDIWRVYGFADIAAKCAIDIKIPKAI
ENHSSOPVTSELSASVSPSLRIRMRVLVHOGLEKPTGADGTNTPLSR
RMATIKLHGKDLMAPODNICHSQILNMAACDARVYRPAACGGIFDVAIVYD
GGGGEITGKILYKEPPIKGFNPDIPIHIEVAQVLDGVNTEGMPDSIPASVITK
WVLDWGKDCIKILKNGKEAVLPNIGVLIIECVIGKRTMTAEERDQLEHVRQ
LDVMVMTSTGKERTLKEPFLVTEAGFARYEVRDQVSLIAYRS"
complement(52130..52610)
/number=1
complement(52611..52939)
/gene="T12J5.20"
/number=2
complement(52940..53064)
/number=2
complement(53065..53389)
/gene="T12J5.20"
/number=3
53847..57666
/gene="T12J5.30"
complement(join(55847..56170,56842..57666))
/gene="T12J5.30"
complement(join(55847..56170,56842..57666))
/note="strong similarity to caffeic acid
O-methyltransferase, Pinus taeda, gb:U39301
contains EST gb:N96576, W43582"
/codon_start=1
/product="O-methyltransferase-like protein"
/protein_id="CAB36723.1"
/db_xref="GI:4455342"
translation="MSSDOLSKFLDRNKMEDNKKRVLDEEAKASLDIWKIYFGEADIA
AAKCAIDIKIPKAIENHSSOPVTSELSASVSPSLRIRMRVLVHOGLEKPTGADGTNTPLSR
DGLATGVNMTSLRRMTTRDQKSLAPVLEFETTPMLARMLSLSVSSPVAGSGP
PPPIAVHCKDVMSRQDNPI LSDMINEMACDARVYRPAVAGCHGLEDFGDTTVAIVDVG
GTGETMGMLVKEPPIKGFNPDIPIHIEVAEVLGAVNTEGMPDSIPACDAIFIKW

VLHWDGDKCIKILKNCKEAVPPNIGKVLIVESVIGENKKTWIVDERDEKLEHVRML
DMVMAHTSTGTEKEMDFYLKEAGFARVERVIDDVSOLIAVRS"
exon complement(55847. .56170)
/gene="t12j5.30"
/number=1
complement(56171. .56841)
/number=1
complement(56842. .57666)
/gene="t12j5.30"
/number=2
61697. .62188
/gene="t12j5.40"
61697. .62188
/gene="t12j5.40"
/number=1
61697. .62188
/gene="t12j5.40"
/note="similarity to predicted protein, Arabidopsis
thaliana"
/codon_start=1
/product="putative protein"
/protein_id="CAB36724.1"
/db_xref="GI:445343"
/translation="MTWPARRSPQNTKRPVYVHSPNTDVKISTSGSPGSP
NDQGVSNFQHSVAESSYPRSSGFLRNEYSYVQVHDLDRTHDEDEYDEMDGDEK
RRRTFRVSCLEFLVLAFTLECLIMGSKSFATLTKVIDRDLRENFETISV
FKI"
gene complement(joln(63375. .63860,63950. .64081,64151. .64551,
64736. .64869,64975. .65243))
/gene="t12j5.50"
complement(63375. .63860)
/gene="t12j5.50"
/number=1
63375. .65243
/gene="t12j5.50"
complement(joln(63375. .63860,63950. .64081,64151. .64551,
64736. .64869,64975. .65243))
/gene="t12j5.50"
/note="strong similarity to Lily mRNA, Lilium longiflorum,
gd:D21814
contains PTS HPR component phosphorylation sites
signatures, Pts_Hpr_Ser [SOEPMNLNLSGVS.LIG]"
/codon_start=1
/product="amino acid permease-like protein"
/protein_id="CAB36725.1"
/db_xref="GI:445344"
/translation="WSIALGNLFDESQSGSPFLMSPAPSTDPQISGEKNGDGG
RIPVEMLPITESRGVNTATFHLGSGIGQVILPAFAALGMVGTILLTGFPV
WLYTTLVLVQHEAVPGIIRISRYRLIASFGACTIIVIGKSIQOLQIMSDN
TAPLTSVOCFLVESCIIAMISQPMNLNLSGVSIGAFMGIACTIVIMLIPVAPSDR
TQVSYSYANMKSFPVHFNALIGLIALVYRGNLVEIOVLTNSOLKHDGTLPSDSK
NPSCTMRRAVMSIALVAICMFPLTPAVYAYGCKIPATGSPVGNLYKLTIOESKR
AACEITHLFTISCSTYPTNLMPCADNEMVITTKRKPASTITVRMLVSLVCEFT
IAGVFPPLPYLAVLIGALALVLTFTYPCFMWISIKKPKRSPMVLNVLVGLASLS
VLLVASAMRLAQKGIHANFFRP"
intron complement(63861. .63949)
/number=1
complement(63950. .64081)
/gene="t12j5.50"
/number=2
complement(64082. .64150)
/number=2
complement(64151. .64551)
/gene="t12j5.50"
/number=3
complement(64552. .64735)
/number=3
complement(64736. .64869)
/gene="t12j5.50"
/number=4
complement(64870. .64974)
/number=4
complement(64975. .65243)
exon

Query Match 0.7%: Score 21; DB 8; Length 84499;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 626 caaatcgtctctctcccg 646
|||||
Db 44958 CAAATTCGTCTCTCTCCG 44978
RESULT 34
ATF23E12
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Eudicotyledons; Core eudicots;
Rosidae; Eurosids II; Brassicales; Brassicaceae; Arabidopsi
1 (bases 1 to 86710)
Bevan,M., Hilbert,H., Braun,M., Holzer,E., Brandt,A.,
Duesterhoeft,A., Hehseisel,J., Mewes,H.W., Mayer,K.F.X. and
Schneller,C.
Unpublished
2 (bases 1 to 86710)
EU Arabidopsis sequencing project.
Direct Submission
Submitted (01-APR-1999) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG. E-mail:
Schneller@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.bevan@bbsrc.ac.uk
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>.
FEATURES
source
gene
exon
CDS
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
173. .1890
/gene="F23E12.10"
173. .374
/gene="F23E12.10"
/number=1
joln<173. .374,456. .919,1089. .1610,1678. .1890)
/gene="F23E12.10"
/codon_start=1
/product="hypothetical protein (fragment)"
/protein_id="CAA18726.1"
/db_xref="GI:3080407"
/db_xref="SPTREMBL:O65486"
/translation="ENAAVAGCFRAVESYLMPSSTSDSTSLPNTTSWYLSAVTASV
VSEIGLSEPAKVPDYDRSPGLPEVLILALGLVSLASRCTSSWTSVADSLNK
DAGIKPAVEPAKGVGLVGIILAVPEVLVWGNVDILKRPFKGASADLLDLVA
VKIATAMCRASGLVGYVAPSLFTFGAGMAYGKFIQALAKQNDNFNSILEVASPO
AYGLVGMATLAVGCOVPLTAVLLFELTODYRIYPLIGAVGSMSTISGGSKQET
RETKETRRKKSQEAOSLTSSDSESTNNLCVESSTCIDDSLNSEELPKRITPYSEA
MTRFRATVMSTSLBALTRMLTEKOSCALIVDPNIFLIGITLSDIOEFSARKEGN
NRPKDIFVNDICSRSGKCKVPTVPMDDLAAOTINMKHELSHVAAYVSGSIDAPRI
HPVGLDRECIITLRR"
intron
exon
/gene="F23E12.10"
/number=1
456. .919
/gene="F23E12.10"
/number=2

```

intron      920..1088
            /gene="F23E12.10"
            /number=2
exon        1089..1610
            /gene="F23E12.10"
            /number=3
intron      1611..1677
            /gene="F23E12.10"
            /number=3
exon        1678..1890
            /gene="F23E12.10"
            /number=4
exon        3124..3241
            /gene="F23E12.20"
            /number=1
gene        3124..4428
            /gene="F23E12.20"
            /join(3124..3241,3323..3465,3595..3696,3945..4107,
            4211..4428)
            /note="Similarity to dihydroflavonol-4-reductase, Gerbera
            x sp., PIR2:S5189"
            /codon_start=1
            /product="putative protein"
            /protein_id="CA18727.1"
            /db_xref="GI:3080408"
            /db_xref="SPTREMBL:O65487"
            /translation="MDQAKGRVCYTGASGLASMLVKRLLLEGVEVIGTVDPGNEKK
            LAHLMKEGAKERLRLYKADLMEGSGEDNAIMGCGVHTASPEILRPATIGTLNVL
            SCRNPSLKRIVLTSSSTVYALSKTLAEQAAKSESGCIDLYTVLPSPLVGSPLP
            PDLCSADVDVGLKGETEKFQMGVGHIDVAPRHIVFEHEAAGKVICSSNV
            ISLEIVSFLSARVPSLPIPKRPDPLN"
            /number=1
intron      3242..3322
            /gene="F23E12.20"
            /number=1
exon        3323..3465
            /gene="F23E12.20"
            /number=2
intron      3466..3594
            /gene="F23E12.20"
            /number=2
exon        3595..3696
            /gene="F23E12.20"
            /number=3
intron      3697..3944
            /gene="F23E12.20"
            /number=3
exon        3945..4107
            /gene="F23E12.20"
            /number=4
intron      4108..4210
            /gene="F23E12.20"
            /number=4
exon        4211..4428
            /gene="F23E12.20"
            /number=5
gene        complement(join(4939..5013,5109..5195,5307..5342,
            5451..5651,5725..5811,6161..6163))
            /gene="F23E12.30"
            /number=1
gene        4939..6163
            /gene="F23E12.30"
            /complement(join(4939..5013,5109..5195,5307..5342,
            5451..5651,5725..5811,6161..6163))
            /gene="F23E12.30"
            /note="Contains Clathrin adaptor complexes small chain
            signature [VVKRYASLYF]
            contains EST gb:744219"
            /codon_start=1
            /product="clathrin assembly protein AP19 homolog"
            /protein_id="CA18728.1"
            /db_xref="GI:3080409"
            /db_xref="SPTREMBL:O23685"
            /translation="MIFVLLVSRQGVRLTKWSPYAOKERKSVIRELSGVILNRCP

```

```

exon        complement(4939..5013)
            /gene="F23E12.30"
            /number=1
intron      complement(5014..5108)
            /number=1
exon        complement(5109..5195)
            /gene="F23E12.30"
            /number=2
intron      complement(5196..5306)
            /number=2
exon        complement(5307..5342)
            /gene="F23E12.30"
            /number=3
intron      complement(5343..5450)
            /number=3
exon        complement(5451..5651)
            /gene="F23E12.30"
            /number=4
intron      complement(5652..5724)
            /number=4
exon        complement(5725..5811)
            /gene="F23E12.30"
            /number=5
intron      complement(5812..6160)
            /number=5
exon        complement(6161..6163)
            /gene="F23E12.30"
            /number=6
exon        complement(7352..7582)
            /gene="F23E12.40"
            /number=1
gene        7352..7582
            /gene="F23E12.40"
            /complement(7352..7582)
            /gene="F23E12.40"
            /complement(7352..7582)
            /gene="F23E12.40"
            /codon_start=1
            /product="hypothetical protein"
            /protein_id="CA18729.1"
            /db_xref="GI:3080410"
            /db_xref="SPTREMBL:O65488"
            /translation="MPLVRKNGVMRIGSPMLHSENLVGDGVEAPPGRDEIGEY
            GGETNVVSESGGEGGAQDAKTPGEADRI"
            /number=1
gene        8233..9045
            /gene="F23E12.50"
            /complement(8233..9045)
            /gene="F23E12.50"
            /complement(8233..9045)
            /gene="F23E12.50"
            /number=1
CDS         complement(8233..9045)
            /gene="F23E12.50"
            /note="Similarity to DNA-binding protein PDI, Pisum
            sativum"
            /product="putative protein"
            /protein_id="CA18730.1"
            /db_xref="GI:3080411"
            /db_xref="SPTREMBL:O65489"
            /translation="MELNRSEADEAKAETTPTGATSSATASGSSGRPRPGAGSK
            NKRPPTITRDSNVLRSHLEVTSQSDISEAVSTVATRRGCGCTISGAVTNTV
            IROPAPAGGCVITLHGRFDLSITGRLALPPAPPGAGGLTVIAGGQGVVGGNVAG

```

```

Query Match      0.7%; Score 21; DB 8; Length 86710;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
07 626 caaatgctctctcccg 646
|||||

```

Db 11744 CAAATTCGTCCTCCTCCG 11764

```
RESULT 35
AC004907 130117 bp DNA PRI 21-AUG-1999
DEFINITION Homo sapiens clone DJ0853H20, complete sequence.
ACCESSION AC004907
VERSION AC004907.2 GI:5757545
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 130117)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 130117)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
3 (bases 1 to 130117)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (21-AUG-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Aug 21, 1999 this sequence version replaced gi:3213090.
FEATURES
    source          1..130117
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="DJ0853H20"
BASE COUNT 43826 a 22237 c 22107 g 41947 t
ORIGIN
Query Match 0.7%; Score 21; DB 9; Length 130117;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2144 cctcagttgtgctcagcac 2164
Db 28683 CCTCAGTTGTGCTCAGCAC 28703
RESULT 36
AC079351 157308 bp DNA HTG 03-FEB-2001
DEFINITION Homo sapiens chromosome 7 clone RP11-828B13, WORKING DRAFT
ACCESSION AC079351
VERSION AC079351.3 GI:12658193
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 157308)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 157308)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (28-AUG-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Feb 3, 2001 this sequence version replaced gi:11415227.
```

```
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0828B13
----- Summary Statistics -----
Sequencing vector: MJ3: 93%
Sequencing vector: plasmid: 6%
Chemistry: Dye-Primer ET; 93% of reads
Chemistry: Dye-terminator Big Dye; 6% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 151471 bases at least Q40
Consensus quality: 153455 bases at least Q30
Consensus quality: 154315 bases at least Q20
Insert size: 16200; agarose-fp
Insert size: 156208; sum-of-ctrls
Quality coverage: 5.45 in Q20 bases; agarose-fp
Quality coverage: 5.28 in Q20 bases; sum-of-ctrls
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 6906: contig of 6906 bp in length
6907 7006: gap of unknown length
7007 16304: contig of 9298 bp in length
16305 16404: gap of unknown length
16405 27392: contig of 10988 bp in length
27393 27492: gap of unknown length
27493 38821: contig of 11329 bp in length
38822 38921: gap of unknown length
38922 53583: contig of 14662 bp in length
53584 53683: gap of unknown length
53684 71646: gap of 17963 bp in length
71647 71746: gap of unknown length
71747 96546: contig of 24800 bp in length
96547 96646: gap of unknown length
96647 119634: contig of 22988 bp in length
119635 119734: gap of unknown length
119735 146540: contig of 26806 bp in length
146541 146640: gap of unknown length
146641 150320: contig of 3680 bp in length
150321 150420: gap of unknown length
150421 153662: contig of 3242 bp in length
153663 153762: gap of unknown length
153763 157308: contig of 3546 bp in length.
FEATURES
    source          1..157308
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="7"
                    /clone="RP11-828B13"
                    1..6906
                    /note="assembly_name:Contig10"
                    7007..16304
                    /note="assembly_name:Contig11"
                    16405..27392
                    /note="assembly_name:Contig12"
                    27493..38821
                    /note="assembly_name:Contig13"
                    38922..53583
                    /note="assembly_name:Contig14
                    clone_end:T7
                    vector_side:left"
                    53684..71646
                    /note="assembly_name:Contig15"
                    71747..96546
                    /note="assembly_name:Contig16"
```

```

misc_feature      96647..119634
                  /note="assembly_name:Contig17"
misc_feature      119735..146540
                  /note="assembly_name:Contig18"
misc_feature      146641..150320
                  /note="assembly_name:Contig7"
misc_feature      150421..153662
                  /note="assembly_name:Contig8"
misc_feature      153763..157308
                  /note="assembly_name:Contig9
                  clone_end:SP6
                  vector_side:right"

BASE COUNT      52313 a 26687 c 27184 g 50018 t 1106 others
ORIGIN

Query Match      0.7%; Score 21; DB 2; Length 157308;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2144 cctcagttgtcctcagcac 2164
      |||
Db 24594 CCTCAGTTGTCTCTCAGCAC 24614

RESULT 37
AC024590/c      183839 bp      DNA      HTG      25-APR-2001
LOCUS      Homo sapiens chromosome 16 clone RP11-487C14, WORKING DRAFT
DEFINITION      AC024590
ACCESSION      AC024590
VERSION      AC024590.4 GI:13786378
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 183839)
AUTHORS      DOE Joint Genome Institute.
TITLE      Sequencing of Human Chromosome 16
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 183839)
AUTHORS      DOE Joint Genome Institute.
TITLE      Direct Submission
JOURNAL      Submitted (29-FEB-2000) Production Sequencing Facility, DOE Joint
AUTHORS      Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
TITLE      On Apr 25, 2001 this sequence version replaced gi:1954669.
JOURNAL      -----Genome Center
AUTHORS      Center: Joint Genome Institute
TITLE      Center Code: JGI
JOURNAL      Web site: http://www.jgi.doe.gov
AUTHORS      -----
TITLE      Project Information
JOURNAL      Center Project Name: 593435
AUTHORS      Center clone name: RPCI-11_487C14
JOURNAL      -----
TITLE      Summary Statistics
JOURNAL      Consensus quality: 172462 bases at least Q40
AUTHORS      Consensus quality: 177746 bases at least Q30
JOURNAL      Consensus quality: 178816 bases at least Q20
AUTHORS      Estimated insert size: 163610; agarose-ftp estimation
JOURNAL      Quality coverage: 14.49 in Q20 bases; agarose-ftp estimation
AUTHORS      Quality coverage: 12.99 in Q20 bases; sum-of-coverage estimation
JOURNAL      NOTE: This is a 'working draft' sequence. It currently
AUTHORS      consists of 14 contigs. The true order of the pieces
JOURNAL      is not known and their order in this sequence record is
AUTHORS      arbitrary. Gaps between the contigs are represented as
AUTHORS      runs of N, but the exact sizes of the gaps are unknown.
AUTHORS      This record will be updated with the finished sequence
AUTHORS      as soon as it is available and the accession number will
AUTHORS      be preserved.
AUTHORS      1
AUTHORS      1928: contig of 1928 bp in length

```

```

FEATURES
source
1..183839
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-487C14"
/clone_lib="RPCI human BAC library 11"
BASE COUNT      50219 a 37425 c 39803 g 55084 t 1308 others
ORIGIN

Query Match      0.7%; Score 21; DB 2; Length 183839;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2012 caggactctgaggtcacact 2032
      |||
Db 42601 CAGGACTCTGAGGTCACACT 42581

RESULT 38
AC013553/c      184864 bp      DNA      HTG      13-DEC-2000
LOCUS      Homo sapiens chromosome 15 clone RP11-325L12 map 15, *** SEQUENCING
DEFINITION      AC013553
ACCESSION      AC013553
VERSION      AC013553.14 GI:11693406
KEYWORDS      HTG; HTGS_PHASE2.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 184864)
AUTHORS      Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE      Homo sapiens chromosome 15, clone RP11-325L12
JOURNAL      Unpublished
AUTHORS      2 (bases 1 to 184864)
REFERENCE      Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
AUTHORS      Baidwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Boukhalter, B.,
AUTHORS      Brown, A., Castle, A., Colangelo, M., Collins, S., Collamore, A.,
AUTHORS      Cooke, P., DeArrellano, K., Dewar, K., Domino, R., Donelan, L., Doyle, M.,
AUTHORS      Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,
AUTHORS      Galand, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
AUTHORS      Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
AUTHORS      Lebeck, J., Liu, C., Locke, K., Macdonald, P., Marquis, N.,
AUTHORS      McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Melidrim, J.,
AUTHORS      Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,

```


Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessier,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (13-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Dec 13, 2000 this sequence version replaced gi:9280765.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green,P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

COMMENT

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
Center project name: L2203
Center clone name: 325_L_12

* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 184864: contig of 184864 bp in length.
Location/Qualifiers
1. 184864
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-325L12"
/clone_lib="RPC1-11 Human Male BAC"

FEATURES

source

BASE COUNT 51274 a 42968 c 41143 g 49479 t
ORIGIN

Query Match 0.7%; Score 21; DB 2; Length 184864;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 294 caaagcttttcacagaca 314
|||||
DB 46943 CAAAGCTTTTCATCAGACA 46923

RESULT 39

AC080090

AC080090 194874 bp DNA HTG 26-NOV-2000
Homo sapiens chromosome 7 clone RP11-79616, WORKING DRAFT SEQUENCE,
29 unordered pieces.

ACCESSION

AC080090

AC080090.3 GI:11345029
HTG: HTGS_PHASE1; HTGS_DRAFT.

KEYWORDS

HTG: HTGS_PHASE1; HTGS_DRAFT.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 194874)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 194874)
Waterston,R.H.
Direct Submission

REFERENCE

AUTHORS

JOURNAL

Submitted (23-SEP-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

COMMENT

On Nov 26, 2000 this sequence version replaced gi:11276326.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC

Web site: http://genome.wustl.edu/gsc/index.shtml

----- Project Information -----
Center project name: H NH0796106

----- Summary Statistics -----
Sequencing vector: M13; 100%

Sequencing vector: plasmid; 0%

Chemistry: Dye-terminator Big Dye; 0% of reads

Chemistry: Dye-terminator Big Dye; 0% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 178730 bases at least Q40

Consensus quality: 184228 bases at least Q30

Consensus quality: 186326 bases at least Q20

Insert size: 191473; sum-of-contigs

Insert size: 191473; sum-of-contigs

Quality coverage: 3.95 in Q20 bases; agarose-fp

Quality coverage: 4.25 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 601: contig of 601 bp in length
1 701: gap of unknown length
* 602 701: gap of unknown length
* 702 2533: contig of 1832 bp in length
* 2534 2633: gap of unknown length
* 2634 3924: contig of 1291 bp in length
* 3925 4024: gap of unknown length
* 4025 5277: contig of 1252 bp in length
* 5277 5376: gap of unknown length
* 5377 6813: contig of 1436 bp in length
* 6813 6912: gap of unknown length
* 6913 9058: contig of 2146 bp in length
* 9059 9158: gap of unknown length
* 9159 11288: contig of 2130 bp in length
* 11289 11389: gap of unknown length
* 11390 13955: contig of 2567 bp in length
* 13956 14055: gap of unknown length
* 14056 16269: contig of 2214 bp in length
* 16270 16369: gap of unknown length
* 16370 19134: contig of 2765 bp in length
* 19135 19234: gap of unknown length
* 19235 22939: contig of 3705 bp in length
* 22940 23039: gap of unknown length
* 23040 26095: contig of 3055 bp in length
* 26096 26195: gap of unknown length
* 26196 29786: contig of 3591 bp in length
* 29787 29886: gap of unknown length
* 29887 32815: contig of 2928 bp in length
* 32816 32914: gap of unknown length
* 32915 37640: contig of 4723 bp in length
* 37641 37739: gap of unknown length
* 37740 41000: contig of 3261 bp in length
* 41001 41100: gap of unknown length
* 41101 46784: contig of 5684 bp in length
* 46785 46884: gap of unknown length
* 46885 54478: contig of 7594 bp in length
* 54479 54578: gap of unknown length
* 54579 61716: contig of 7138 bp in length
* 61717 61816: gap of unknown length
* 61817 68787: contig of 6970 bp in length
* 68788 68887: gap of unknown length
* 68888 76504: contig of 7618 bp in length
* 76505 86157: gap of unknown length
* 86158 86157: contig of 9553 bp in length

```

*      86158      86257: gap of unknown length
*      86258      95154: contig of 8897 bp in length
*      95155      95254: gap of unknown length
*      95255      105637: contig of 10383 bp in length
*      105638     105737: gap of unknown length
*      105738     119880: contig of 14143 bp in length
*      119881     119980: gap of unknown length
*      131203     131302: contig of 11222 bp in length
*      131303     144326: contig of 13024 bp in length
*      144327     144426: gap of unknown length
*      144427     164332: contig of 19906 bp in length
*      164333     164432: gap of unknown length
*      164433     194874: contig of 30442 bp in length.
*
Location/Qualifiers
1. 194874
/organism="Homo Sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/clone="RP11-79616"
1. .601

```

```

misc_feature /note="assembly_name:Contig10"
702. 2533
/note="assembly_name:Contig19"
2634. 3924
/note="assembly_name:Contig11"
4025. 5276
/note="assembly_name:Contig14"
5377. 6812
/note="assembly_name:Contig16"
6913. 9058
/note="assembly_name:Contig17"
9159. 11288
/note="assembly_name:Contig18"
11389. 13955
/note="assembly_name:Contig20"
14056. 16269
/note="assembly_name:Contig21"
16370. 19134
/note="assembly_name:Contig22"
19235. 22939
/note="assembly_name:Contig23"
23040. 26095
/note="assembly_name:Contig24"
26196. 29786
/note="assembly_name:Contig25"
29887. 32814
/note="assembly_name:Contig26"
32915. 37639
/note="assembly_name:Contig27"
37740. 41000
/note="assembly_name:Contig28"
41101. 46784
/note="assembly_name:Contig29"
46885. 54478
/note="assembly_name:Contig30"
54579. 61716
/note="assembly_name:Contig31"
61817. 68786
/note="assembly_name:Contig32"
68887. 76504
/note="assembly_name:Contig33"
76605. 86157
/note="assembly_name:Contig34"
86258. 95154
/note="assembly_name:Contig35"
95255. 105637
/note="assembly_name:Contig36"
105738. 119880
/note="assembly_name:Contig37"
119981. 131202
/note="assembly_name:Contig38"
131303. 144326
misc_feature

```

```

misc_feature /note="assembly_name:Contig39"
14427. 164332
/note="assembly_name:Contig40"
164433. 194874
/note="assembly_name:Contig41"
194874. 2806 others
BASE COUNT 63657 a 32523 c 33628 g 62260 t
ORIGIN

```

```

Query Match 0.7% Score 21: DB 2: Length 194874;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 2144 cctcattgtgcctcagcac 2164
|||||
Db 52703 CCTCAGTTGGTCCTCAGCAC 52723

```

```

RESULT 40
ATCHRIV82 LOCUS
DEFINITION Arabidopsis thaliana DNA chromosome 4, contig fragment No. 82.
ACCESSION AL161586
VERSION AL161586.2 GI:7270418
KEYWORDS
SOURCE ORGANISM
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 29459 to 132608; 129923 to 195165)
Murphy,G., Ridley,P., Hudson,S., Mewes,H.W., Lemcke,K. and
Mayer,K.F.X.
Unpublished
2 (bases 98968 to 179313)
Purnelle,B., Boutry,M., Goffeau,A., Mewes,H.W., Lemcke,K. and
Mayer,K.F.X.
Unpublished
3 (bases 1 to 44270)
Terryn,N., Ardijns,W., Buysseart,C., Dasseville,R., De Clerck,R.,
De Keyser,A., Neyt,P., Rouze,P., Van Den Daele,H., Villarroel,R.,
Gielens,J., Van Montagu,M., Mewes,H.W., Lemcke,K. and Mayer,K.F.X.
Unpublished
4 (bases 1 to 195165)
EU Arabidopsis sequencing project.
Direct Submission
Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG. E-mail:
lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de,project
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.bevan@jbsrc.ac.uk

```

```

COMMENT
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
this fragment has an overlap with ATCHRIV81 at the 5' end and an
overlap with ATCHRIV83 at the 3' end.

```

```

FEATURES
Source
1. 195165
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
8478. 8759
/gene="AT4g34680"
/number=1
8478. 9388
/gene="AT4g34680"
join(8478..8759,8861..9388)
/gene="AT4g34680"
/note="contains EST gb:AI994545.1, R30028"
/codon_start=1
/product="GATA transcription factor 3"

```

/protein_id="CAB80185.1"
 /db_xref="GI:7270419"
 /translation="MEIWTAEARLAKSLRGESTISLKHQIVASEDLSPRTSSLPEDFS
 VECFLDFSGCKEEFEVYSSVSSOEOEHDCVSSQPCIFDLPSPDPDVEELE
 WYSRVVDDCSSPEVSLTLTQTHKTPSPSRILEVKKPRTRSRNSLTGSRWPLVSTNHQ
 HAATQDLRRKKQETLVFQRCSHCGTNTPDWRTGVPKTLCAACGAFKFSGLCP
 EYRPADSPFSSNEIHSNLHRKVELRKSEIGEETGEASTKSDPVKFGSKW"
 8760..8860
 /gene="AT4g34680"
 /number=1
 8861..9388
 /gene="AT4g34680"
 /number=2
 10241..10875
 /gene="AT4g34690"
 /complement(10241..10855)
 /number=1
 /complement(join(10241..10855,10867..10875))
 /gene="AT4g34690"
 /complement(join(10241..10855,10867..10875))
 /gene="AT4g34690"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="CAB80186.1"
 /db_xref="GI:7270420"
 /translation="MDISDDSGEKKKNDGSGDKKSDGSSNNDDDDYTPNDPOS
 TWFKYFLATRLHYGVKRCSESVKRWKFLKLTINISQAKLQDLAKAKKEHOTF
 VEYTLGKASRDSQTFYSLTWCFGSJLTCSFVWGLMNTSKTRSLADLWLEINKK
 SVVIHEPKKMSAKNALYMRDLDEITTKAKONQEDLQKIVALVSKK"
 /complement(10856..10866)
 /number=1
 /complement(10867..10875)
 /gene="AT4g34690"
 /number=2
 11652..13140
 /gene="AT4g34700"
 /join(11652..11783,12719..12800,12946..13001,13099..13140)
 /gene="AT4g34700"
 /note="contains EST gb:T04342, T13747, AA040992,
 A1992559.1, Z33673, H36271, T43433, T43941"
 /codon_start=1
 /product="putative protein"
 /protein_id="CAB80187.1"
 /db_xref="GI:7270421"
 /translation="MSGVSTAYFARRAAQKRVRLILRRALKDILNNAVHHIRYRD
 EDVDRIDKLIAGEAEYKMKRHPDPIVPMAGSGKFCENPTPPAGIETIVNYGLEDN
 P"
 11652..11783
 /gene="AT4g34700"
 /number=1
 11784..12718
 /gene="AT4g34700"
 /number=1
 12719..12800
 /gene="AT4g34700"
 /number=2
 12801..12945
 /gene="AT4g34700"
 /number=2
 12946..13001
 /gene="AT4g34700"
 /number=3
 13002..13098
 /gene="AT4g34700"
 /number=3
 13099..13140
 /gene="AT4g34700"
 /number=4
 /complement(15093..17228)
 /gene="AT4g34710"
 /complement(15093..17228)
 /gene="AT4g34710"
 /translation="MAPVNLPTNLVSRRTVRGMANPBRVKVAKOIMRELSMDLITDT

/note="Contains Oru/DAP/Arg decarboxylases family 2
 signatures AA144-162.Oru/DAP/Arg decarboxylases family 2
 signatures AA325-338
 contains EST gb:W43783, Z33699, Z33688, T46784,
 A1998693.1, AA404845"
 /codon_start=1
 /product="arginine decarboxylase SPE2"
 /protein_id="CAB80188.1"
 /db_xref="GI:7270422"
 /translation="MPALACVPTSTVPAPYAFSDTAGVDFIAPSSFTSANVYVDRMS
 SLSSSLVRIDGWGAPYFIANSSGNISVPRHGETLPDQIDILKLVKVTGPKSSGGL
 GLQPLIVAFEPDLNRLCLOASDADYVLSQGYDHPVPCNDQRFVEDIVK
 FGSSRFEGLEASKEPILAMSCLEKSPDAFLVNGFQDAEYISLALGRKLTANTV
 IVLBOEELDLVIELTSOKMNVPRVIGLRKLTKSGHGRGSGRGLTQTQIVR
 VYRKLRQSGMDCLDLHLHISQITPSTSLSDGYAEAAQITCEIVRGAAHKVYDIDG
 GGLGIDYDGSKGSSEDSLVASISLEIYAEVAVSVVCDRSSVKHPVICSSEGRAIVS
 HHSLVLIIEAVSADKPMVHQATPGDIOFLLEGEEARANVEDIYAAVMGRDHSCILY
 DQIKORCVGEKEGVLSIEQLASVDGLEWYKAIGASDPVHTYINLSVFTSIPDLW
 GIDQLFPIYPIHKLDORPGARCIISDLTCDSDGKINKEFGESSESLPHELDNGSGR
 YELGMEIGAYEALGVYHNLFGSPSVRVSDGPHSPAVTRAVPDGSADVLRAMO
 HPEELMFOTIKRAEEMHTKGSSEGENEEDEDFNNVAVSLDRSRNMPLYLATEQ
 ASPSNSLSAISNLGFTYCDQEDVYDTISA"
 /complement(15093..17228)
 /gene="AT4g34710"
 /number=1
 15093..17228
 /gene="AT4g34710"
 /complement(join(23001..23133,23231..23516,23868..23943))
 /gene="AT4g34720"
 /complement(23001..23133)
 /gene="AT4g34720"
 /number=1
 23001..23943
 /gene="AT4g34720"
 /complement(join(23001..23133,23231..23516,23868..23943))
 /gene="AT4g34720"
 /note="Contains Prokaryotic membrane lipoprotein lipid
 attachment site AA16-26:Prokaryotic membrane lipoprotein
 lipid attachment site AA91-101
 contains EST gb:T41774, N96995, N38619, Z26112, N38618,
 A1998496.1, A1996086.1, R90087, AA12128, H76763"
 /codon_start=1
 /product="vacuolar H⁺-transporting ATPase 16k chain"
 /protein_id="CAB80189.1"
 /db_xref="GI:7270423"
 /translation="MSTSGDETFAPFPEFLGAAALVSCMAAGTAKSGVAGSAGK
 VWRPELVKSLIVPVMAVGVLGIYLIVIIISTGINPAKSYLFDGYAHLSSGLACG
 LAGLSGMAIGIVGDAGVARNADQPKLEVMILLIIFAFALALYGLIYGIILSRAGQ
 SRAE"
 /complement(23134..23230)
 /number=1
 /complement(23231..23516)
 /gene="AT4g34720"
 /number=2
 /complement(23517..23867)
 /number=2
 /complement(23868..23943)
 /gene="AT4g34720"
 /number=3
 24797..25947
 /gene="AT4g34730"
 /complement(join(24797..24955,25226..25399,25678..25947))
 /gene="AT4g34730"
 /complement(join(24797..24955,25226..25399,25678..25947))
 /gene="AT4g34730"
 /note="similarity to similarity predicted protein,
 Synechocystis sp., PIR2:576285
 contains EST gb:F19817"
 /codon_start=1
 /product="putative protein"
 /protein_id="CAB80190.1"
 /db_xref="GI:7270424"
 /translation="MAPVNLPTNLVSRRTVRGMANPBRVKVAKOIMRELSMDLITDT

VL0HAYLPEALCADRYLSTTTSDVEVSNDLOVMRPNMKLNCNVKVVSVFGDD

Query Match 0.7%; Score 21; DB 8; Length 195165;

Best Local Similarity 100.0%; Pred. No. 21;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 626 caattgctctctctctccg 646

Db 174880 CAAATCTCTCTCTCCG 174900

RESULT 41

AC025566/c

LOCUS

DEFINITION

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

(see http://www.hgsc.bcm.tmc.edu/docs/genbank-draft_data.html).

NOTE: This is a 'working draft' sequence. It currently consists of 40 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1	17628:	contig of 17628 bp in length
*	17728:	gap of unknown length
*	17729:	contig of 12923 bp in length
*	30651:	gap of unknown length
*	30751:	gap of unknown length
*	30752:	contig of 9059 bp in length
*	39810:	gap of unknown length
*	39811:	gap of unknown length
*	39911:	contig of 7994 bp in length
*	47904:	gap of unknown length
*	47905:	contig of 806 bp in length
*	48005:	contig of unknown length
*	56311:	gap of unknown length
*	56411:	contig of 7158 bp in length
*	63568:	gap of unknown length
*	63569:	contig of 6656 bp in length
*	70324:	contig of 6656 bp in length
*	70325:	gap of unknown length
*	70424:	gap of unknown length
*	77997:	contig of 7573 bp in length
*	78097:	gap of unknown length
*	85133:	contig of 7036 bp in length
*	85134:	gap of unknown length
*	85234:	contig of 5950 bp in length
*	91183:	contig of 5950 bp in length
*	91283:	gap of unknown length
*	95657:	contig of 4384 bp in length
*	95658:	contig of unknown length
*	95767:	gap of unknown length
*	101736:	contig of 5969 bp in length
*	101737:	gap of unknown length
*	101837:	contig of 6802 bp in length
*	108639:	gap of unknown length
*	108738:	gap of unknown length
*	114324:	contig of 5586 bp in length
*	114424:	gap of unknown length
*	118727:	contig of 4303 bp in length
*	118728:	gap of unknown length
*	118827:	contig of 3942 bp in length
*	122769:	contig of 3942 bp in length
*	122869:	gap of unknown length
*	129332:	contig of 6463 bp in length
*	129333:	contig of unknown length
*	129433:	gap of unknown length
*	129434:	contig of 4734 bp in length
*	134167:	gap of unknown length
*	134267:	contig of 5058 bp in length
*	139325:	gap of unknown length
*	139424:	gap of unknown length
*	144289:	contig of 4865 bp in length
*	144389:	gap of unknown length
*	144390:	contig of 4773 bp in length
*	149163:	gap of unknown length
*	149164:	contig of 3669 bp in length
*	152931:	contig of 3669 bp in length
*	152932:	gap of unknown length
*	153031:	contig of 4633 bp in length
*	157664:	contig of unknown length
*	157665:	gap of unknown length
*	157765:	contig of 5662 bp in length
*	163426:	gap of unknown length
*	163526:	contig of 3277 bp in length
*	166803:	contig of 3277 bp in length
*	166903:	gap of unknown length
*	169178:	contig of 2275 bp in length
*	169179:	gap of unknown length
*	169279:	contig of 3370 bp in length
*	172648:	gap of unknown length
*	172748:	contig of 2431 bp in length
*	175179:	gap of unknown length
*	175279:	contig of 2093 bp in length
*	177372:	gap of unknown length
*	177373:	contig of 1864 bp in length
*	179336:	gap of unknown length
*	179337:	contig of 1092 bp in length
*	180528:	gap of unknown length
*	180529:	contig of 1601 bp in length
*	182228:	gap of unknown length
*	182329:	contig of 1601 bp in length
*	182330:	gap of unknown length

* NOTE: Estimated insert size may differ from sequence length

* 182330 184954: contig of 2625 bp in length
* 184955 185054: gap of unknown length
* 185055 186495: contig of 1441 bp in length
* 186496 186595: gap of unknown length
* 186596 188317: contig of 1722 bp in length
* 188318 188417: gap of unknown length
* 188418 189682: contig of 1265 bp in length
* 189683 189782: gap of unknown length
* 189783 191715: contig of 1933 bp in length
* 191716 191815: gap of unknown length
* 191816 192922: contig of 1107 bp in length
* 192923 193022: gap of unknown length
* 193023 194102: contig of 1080 bp in length
* 194103 194202: gap of unknown length
* 194203 195217: contig of 1015 bp in length.
Location/Qualifiers
1. 195217
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-50102"
BASE COUNT 61010 a 36076 c 34593 g 59532 t 4006 others
ORIGIN
Query Match 0.7%; Score 21; DB 2; Length 195217;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1146 atacagaaaaaacacacaa 1166
|||||
Db 188931 ATACAGAAAAACACACAAA 188911
RESULT 42
ATCHRIV83/c
LOCUS ATCHRIV83 197859 bp DNA PLN 16-MAR-2000
DEFINITION Arabidopsis thaliana DNA chromosome 4, contig fragment No. 83.
ACCESSION AL161587
VERSION AL161587.2 GI:7270470
KEYWORDS
SOURCE
ORGANISM
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 16641 to 103350)
Hilbert,H., Braun,M., Holzer,E., Brandt,A., Duesterhoeft,A.,
Mewes,H.W., Lemcke,K. and Mayer,K.F.X.
Unpublished
2 (bases 98352 to 125759)
Robben,J., Grymonprez,B., Volckaert,G., Mewes,H.W., Lemcke,K. and
Mayer,K.F.X.
Unpublished
3 (bases 120761 to 197859)
Rose,M., Hempel,S., Entlan,K.-D., Mewes,H.W., Lemcke,K. and
Mayer,K.F.X.
Unpublished
4 (bases 1 to 24256)
Murphy,G., Ridley,P., Hudson,S., Mewes,H.W., Lemcke,K. and
Mayer,K.F.X.
Unpublished
5 (bases 1 to 197859)
EU Arabidopsis sequencing project.
Direct Submission
Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
lemcke@mps.biochem.mpg.de, mayer@mps.biochem.mpg.de, Project
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.bevan@bsrc.ac.uk
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4

and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>
this fragment has an overlap with ATCHRIV82 at the 5' end and an
overlap with ATCHRIV84 at the 3' end.
FEATURES
source
1. 197859
Location/Qualifiers
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
11339..12705
/gene="AT4g35190"
join(11339..11494,11536..11664,11853..11952,12119..12189,
12314..12457,12499..12705)
/gene="AT4g35190"
note="similarity to ctf, Mycobacterium leprae, gb:U15180
contains EST gb:T45691, T22640, T75954"
/codon_start=1
/product="putative protein"
/protein_id="CAB80236.1"
/db_xref="GI:7270471"
/translation="MEIVKSRKRCVRCGSSGKRECSAATDLAQLVRLCLNLN
ESLENKQVTRRLNLYGGSGIGLGLVSAVHEAGVAGVAGIYDLFTLTGFTYG
EVIAYADMERKAEMARHSDCFIALPGVGTLELLEIYAMAQLGIDHKPVLNVWG
YNYLTFIDKAVDDGFIKPSORHIFVSAFPAKLELVOKLEILMKINKDKKFPESADLL
FPDIOAYKPVNDGVYAKSRWEVEKKVQDPQQQGVVFCSTNSMTETAL"
11339..11494
/gene="AT4g35190"
/number=1
11495..11535
/gene="AT4g35190"
/number=1
11536..11664
/gene="AT4g35190"
/number=2
11665..11852
/gene="AT4g35190"
/number=2
11853..11952
/gene="AT4g35190"
/number=3
11953..12118
/gene="AT4g35190"
/number=3
12119..12189
/gene="AT4g35190"
/number=4
12190..12313
/gene="AT4g35190"
/number=4
12314..12457
/gene="AT4g35190"
/number=5
12458..12498
/gene="AT4g35190"
/number=5
12499..12705
/gene="AT4g35190"
/number=6
13757..14518
/gene="AT4g35200"
complement(13757..14518)
/gene="AT4g35200"
complement(13757..14518)
/gene="AT4g35200"
note="similarity to various predicted proteins,
Arabidopsis thaliana"
/codon_start=1
/product="putative protein"
/protein_id="CAB80237.1"
/db_xref="GI:7270472"
/translation="MAVSFHSNSYPSRQHPQAAVDEQTLRLSSDSASSSSICQK
LSNODLHDSLEKMRILSVTNALSDQIDLEKLDGSLRLDLCNTAKDAISOMKGLM
ETQSLRRKPGDLSEGVKKYLVSRFLKSLQKVIKSLKVCOSKSTNASTLVFCRAE

ATVMALESLFSEFMSGSKACGKMSLVSKMMSQNKVTCBAEANEFTRIDSEFQSEKSLQ
MEDVQNLSECTIQLEDEGIESLSKSLIKYRVSLINI"
complement(13157..14518)
/gene="At4g35200"
/number=1
16043..16795
/gene="At4g35210"
16043..16795
/gene="At4g35210"
/number=1
16043..16795
/gene="At4g35210"
/note="similarity to various predicted proteins,
Arabidopsis thaliana"
/codon_start=1
/product="putative protein"
/protein_id="CAB80238.1"
/db_xref="GI:7270473"
/translation="MAVSFHVRSSTSPSRQHPQAAHYDEQTLRLRSCTASSSTICOR
LSMDQDLHSLERKMLRLSVTQALSDQIEKLKLSIKILDLCSIKDGLSQMKESTIK
EIGSIYRRKRGDLSAEVKKYLAERKFLKSEFVKLSLKTSONKNDALAVGEAEVTV
IALFESLSEFMSGSKACGKMSLVSKMMSQNKVTCBAEANEFTRIDSEFQSEKSLQMED
VONLEICIQLEDEGIESLSKSLIKYRVSLINI"
16640..24256
/note="position 76883-84499 overlaps to BAC clone F23E12,
EMBL acc:AL022604; for sequence analysis please refer to
this accession"
17252..17657
/gene="At4g35220"
/number=1
17252..18625
/gene="At4g35220"
join(17252..17657,17740..17784,18534..18625)
/gene="At4g35220"
/note="similarity to predicted protein, Arabidopsis
thaliana
contains EST gb:FL4002, T46305, T43454, A1996658.1"
/codon_start=1
/product="putative protein"
/protein_id="CAB80239.1"
/db_xref="GI:7270474"
/translation="MAVPLPFLITLISLISLISAGASNAVPSICTAPIDGFTPE
LKIRIRRYGNGKTYDISHRYTPPMSPWDSSEGIQPLMLAASKKNSLANNSEMKIP
THRTGHDSFGHYDKYDAGFDVDSLQVGLALLVDVPPKDKNTITLDEQGIQF
VASSIDSLSQCOIRNCP"
17658..17739
/gene="At4g35220"
/number=1
17740..17784
/gene="At4g35220"
/number=2
17785..18533
/gene="At4g35220"
/number=2
18534..18625
/gene="At4g35220"
/number=3
19941..22657
/gene="At4g35230"
complement(join(19941..20231,20345..20458,20546..20710,
20792..20980,21340..21425,21551..21656,21755..21888,
22078..22213,22304..22657))
/gene="At4g35230"
complement(join(19941..20231,20345..20458,20546..20710,
20792..20980,21340..21425,21551..21656,21755..21888,
22078..22213,22304..22657))
/gene="At4g35230"
/note="similarity to protein kinase APK1, Arabidopsis
thaliana, PIR2:S28615
contains EST gb:A1997799.1, F13911"
/codon_start=1
/product="putative protein"
/protein_id="CAB80240.1"

/db_xref="GI:7270475"
/translation="MCCOSLFGSDNPLRGDGVQPPQLSONNHGATTAADNGSGGAS
GVGGGGGGGJIPSESEFADIKATNNFSDNIVSSGKAPMLVYKGLQNRMTA
VKRETKMAPEEPKQFAEAMGVGLRHNRLINLIGCCDDERLLVKEFNPDLAKH
LFHEMNOTIMAMRLRGVYIAEGLCSREGLRGRLHYDIAVRLPDEDGDLRLSCG
LMKNSRDGKSYSTLATTPEYRLNGVTPESVYTSRGTYLLDLSSKHIPPSALDM
IRGNKILILMDSHLEGFSTEEATVVELLSQCLQEPREPRNKDILVATLAPLQTKS
DVPSYVNLIGRKQCEAPSTPORPLSPGEACSRMDLEAIHQILVMEHYRDEGTNELS
FQEWTOQMCKMDLARKRGDOSFREKDKRTIDCSQMSYGYIHCFCQFIDVGTWVSP
TVFGRSLCYLLCDPPDALRDAAQACVYPDWPTARYMQSVALLAKLNMWTDADMLN
EAOLEEKRRGGRGS"
complement(19941..20231)
/gene="At4g35230"
/number=1
complement(20232..20344)
/number=1
complement(20345..20458)
exon
intron
intron
exon
Query Match
Best Local Similarity 100.0%; Score 21; DB 8; Length 197859;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 626 caaatgctctctccccc 646
Db 91607 CAAATTCGTCCTTCCTCCG 91587
RESULT 43
AC068573/c
LOCUS
DEFINITION
Homo sapiens chromosome 15 clone RP11-540B6 map 15, WORKING DRAFT
SEQUENCE, 12 unordered pieces.
ACCESSION
AC068573
VERSION
AC068573.3 GI:12313774
KEYWORDS
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 198935)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 15, clone RP11-540B6
unpublished
2 (bases 1 to 198935)
Anderson,S., Baldwin,D., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collins,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,D.S.,
Dodgson,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Laroque,K., Lamazares,R., Landers,J., Lehotzky,J.,
Leyland,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Margulis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPeckers,R.,
Meldrum,J., Meneses,L., Minova,T., Miranda,C., Mienga,C., Morrow,D.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisanil,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rotman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessfaye,S., Theodore,D., Tirrell,A., Travers,M., Triggillo,P.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (04-MAY-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jan 19, 2001 this sequence version replaced gi:10280764.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: MIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L10276
 Center clone name: 540_E_6
 ----- Summary Statistics
 Sequencing vector: M13: M77815; 41% of reads
 Sequencing vector: Plasmid; n/a; 59% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 19353 bases at least Q40
 Consensus quality: 195810 bases at least Q30
 Consensus quality: 196898 bases at least Q20
 Insert size: 166000; agarose-fp
 Insert size: 197835; sum-of-contigs
 Quality coverage: 10.6 in Q20 bases; agarose-fp
 Quality coverage: 8.9 in Q20.
 NOTE: This is a 'working draft' sequence. It currently
 * consists of 12 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1
 4883 4982: contig of 4882 bp in length
 4983 5624: contig of 642 bp in length
 5625 5724: gap of 100 bp
 5725 6734: contig of 1010 bp in length
 6735 6834: gap of 100 bp
 6835 8393: contig of 1559 bp in length
 8394 8493: gap of 100 bp
 8494 11316: contig of 2823 bp in length
 11317 11416: gap of 100 bp
 11417 16171: contig of 4755 bp in length
 16172 16271: gap of 100 bp
 16272 30341: contig of 14070 bp in length
 30342 30441: gap of 100 bp
 30442 49473: contig of 19032 bp in length
 49474 49573: gap of 100 bp
 49574 75583: contig of 26010 bp in length
 75584 75683: gap of 100 bp
 75684 158422: contig of 82739 bp in length
 158423 158522: gap of 100 bp
 158523 180063: contig of 21541 bp in length
 180064 180163: gap of 100 bp
 180164 198935: contig of 18772 bp in length.

FEATURES
 source
 1.198935 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="15"
 /map="15"
 /clone_lib="RP11-540E6"
 1.4882 /clone_lib="RP11 Human Male BAC"
 misc_feature
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:left"
 4983 5624
 misc_feature
 /note="assembly_fragment"
 5725 6734
 /note="assembly_fragment"
 6835 8393
 misc_feature
 /note="assembly_fragment"
 8494 11316
 /note="assembly_fragment"
 11417 16171
 /note="assembly_fragment"
 16272 30341
 misc_feature
 /note="assembly_fragment"

misc_feature
 /note="assembly_fragment"
 30442 49473
 /note="assembly_fragment"
 49574 75583
 misc_feature
 /note="assembly_fragment"
 75684 158422
 /note="assembly_fragment"
 158523 180063
 /note="assembly_fragment"
 180164 198935
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:right"
 53896 a 46178 c 45676 g 52083 t 1102 others

BASE COUNT
 ORIGIN

Query Match
 Best local Similarity 100.0%; Pred. No. 21;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 294 caaagcttttcatacagaca 314
 Db 13730 CAAGCTTTTCATCAGACA 13710

RESULT 44
 AL593857/c
 LOCUS
 DEFINITION
 PROGRESS *** in unordered pieces.
 AL593857
 ACCESSION
 AL593857.1 GI:14787341
 VERSION
 HTGS: HTGS-PHASE1; HTGS-DRAFT; HTGS-FULLTOP.
 KEYWORDS
 SOURCE
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 206137)
 SImS.S.
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

COMMENT
 Direct Submission
 Submitted (11-Jul-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 requests: clonequest@sanger.ac.uk
 ----- Genome Center
 Center: Sanger Centre
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquerry@sanger.ac.uk
 ----- Project Information
 Center project name: BM439H2
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Sequencing vector: Plasmid; 108752; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Consensus quality: 204045 bases at least Q40
 Consensus quality: 204532 bases at least Q30
 Consensus quality: 204877 bases at least Q20
 Insert size: 205337; sum-of-contigs
 Insert size: 207948; 2.7% error; agarose-fp
 Quality coverage: 10.60x in Q20 bases; sum-of-contigs Quality
 coverage: 10.47x in Q20 bases; agarose-fp

 NOTE: This is a 'working draft' sequence.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

FEATURES
 source
 1.206137 Location/Qualifiers
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="2"
 /clone_lib="RP23-439H2"

```

misc_feature      /clone.lib="RPCT-23"
                  1..15760
                  /note="assembly_fragment:02571
                  fragment_chain:1
                  clone_end:SP6
                  vector_side:left"
misc_feature      /note="assembly_fragment:00691
                  fragment_chain:1"
                  19857..46285
                  /note="assembly_fragment:04142
                  fragment_chain:1"
misc_feature      46386..146248
                  /note="assembly_fragment:04376
                  fragment_chain:1"
misc_feature      146349..150461
                  /note="assembly_fragment:02422
                  fragment_chain:1"
misc_feature      150562..161265
                  /note="assembly_fragment:04927
                  fragment_chain:1"
misc_feature      161366..165284
                  /note="assembly_fragment:05216
                  fragment_chain:1"
misc_feature      165385..186792
                  /note="assembly_fragment:03899
                  fragment_chain:2"
misc_feature      186893..206137
                  /note="assembly_fragment:00210
                  fragment_chain:2
                  clone_end:T7
                  vector_side:right"
BASE COUNT      57514 a 47684 c 46179 g 53957 t 803.others
ORIGIN

Query Match      0.7%; Score 21; DB 2; Length 206137;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1152 aaaaacaacacacacataaa 1172
      ||||||||||||||||||
Db 63514 AAAAAACAACACACATATAA 63494

RESULT 45
LOCUS      AL513468      211030 bp      DNA      HTG      21-JUL-2001
DEFINITION Mus musculus chromosome X clone RP23-10416. *** SEQUENCING IN
PROGRESS ***, in unordered pieces.
ACCESSION      AL513468
VERSION      1
KEYWORDS      HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 211030)
Direct Submission
Submitted (20-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
On Apr 27, 2001 this sequence version replaced gi:13568261.
----- Genome Center
Center: UK Medical Research Council
Center code: UK-MRC
Web site: http://mrcseq.har.mrc.ac.uk
Contact: mouseseq@har.mrc.ac.uk
----- Project Information
Center project name: bm10416
----- Summary Statistics
Assembly program: XGAP4; version 4.5

```

```

Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 98% of reads
Chemistry: Dye-Primer Big Dye; 1% of reads
Consensus quality: 210707 bases at least Q40
Consensus quality: 210871 bases at least Q30
Consensus quality: 210909 bases at least Q20
Insert size: 210930; sum-of-ctrls
Quality coverage: 8.25x in Q20 bases; sum-of-ctrls quality
coverage: 8.49x in Q20 bases; agarose-tp
-----
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
source      Location/Qualifiers
            1..211030
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /chromosome="X"
            /clone="RP23-10416"
            /clone.lib="RPCT-23"
            1..166819
            /note="assembly_fragment:00499"
            166920..211030
            /note="assembly_fragment:01558"
BASE COUNT      70605 a 40504 c 38146 g 61675 t 100.others
ORIGIN

Query Match      0.7%; Score 21; DB 2; Length 211030;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1152 aaaaacaacacacacataaa 1172
      ||||||||||||||||||
Db 135934 AAAAAACAACACACATATAA 135914

RESULT 46
LOCUS      AC092992      221341 bp      DNA      HTG      09-AUG-2001
DEFINITION Homo sapiens chromosome 3p clone RP11-500K7, WORKING DRAFT
SEQUENCE, 20 unordered pieces.
ACCESSION      AC092992
VERSION      1
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 221341)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Aisbrooks,S.L., Amaralunge,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burke,P., Burrell,C., Burrell,K.L., Byrd,N.C., Carton,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davoy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Dean,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hayes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Homsl,F., Howard,S., Huber,J., Hulik,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,

```


TITLE
JOURNAL
AUTHORS
JOURNAL
COMMENT

Louisegeed, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawlin, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabadi, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogutu, M., Okunolu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojurokan, I., Rolfe, M., Ruiz, S., Saverly, G., Scherer, S., Scott, G., Shen, H., Shoochari, N., Slismon, I., Sodergren, E., Sonalke, T., Sparks, A., Stanek, H., Stone, H., Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, S., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

Unpublished
Direct Submission
2 (bases 1 to 221341)
Worley, K.C.
Direct Submission
Submitted (09-AUG-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

----- Genome Center -----
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

----- Project Information -----
Center project name: HPCU
Center clone name: RP11-500K7

----- Summary Statistics -----
Assembly program: Phrap; version 0.990329
Consensus quality: 212184 bases at least Q40
Consensus quality: 217356 bases at least Q30
Consensus quality: 220392 bases at least Q20
Estimated insert size: 220263; sum-of-coverage estimation
Quality coverage: 0x in Q20 bases; average-tp estimation
Quality coverage: 4.1x in Q20 bases; sum-of-coverage estimation

----- NOTE: Estimated insert size may differ from sequence length -----
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
NOTE: This is a "working draft" sequence. It currently consists of 20 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 22546: contig of 22546 bp in length
22547 22646: gap of unknown length
22647 48410: contig of 25764 bp in length
48411 48510: gap of unknown length
48511 71072: contig of 22562 bp in length
71073 71172: gap of unknown length
71173 91598: contig of 20426 bp in length
91599 91698: gap of unknown length
91699 108834: contig of 17136 bp in length
108835 108935: gap of unknown length
108936 126517: contig of 17583 bp in length
126518 126617: gap of unknown length
126618 139129: contig of 12512 bp in length
139130 139229: gap of unknown length
139230 148601: contig of 9372 bp in length
148602 148701: gap of unknown length
148702 157335: contig of 8634 bp in length
157336 157435: gap of unknown length
157436 166202: contig of 8767 bp in length
166203 166302: gap of unknown length

FEATURES
source 1. 221341
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3q"
/clone="RP11-500K7"

BASE COUNT 70894 a 38966 c 37158 g 72403 t 1920 others

ORIGIN

Query Match 0.7%; Score 21; DB 2; Length 221341;
Best local Similarity 100.0%; Pred. No. 21;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1146 ataccagaaacacacacaa 1166
Db 41072 ATACAGAAACAAACACAAA 41052
|||||
AC020727 221194 bp DNA HTG 10-SEP-2000
LOCUS Homo sapiens chromosome 3 clone RP11-500K7, WORKING DRAFT SEQUENCE,
DEFINITION 20 unordered pieces.
AC020727
AC020727.5 GI:10048064
VERSION HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 221194)
Waterston, R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 221194)
Waterston, R.H.
Direct Submission
Submitted (08-JAN-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
On Sep 10, 2000 this sequence version replaced gi:7232174.

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
----- Project Information -----
Center project name: H.NH0500K07
----- Summary Statistics -----
Sequencing vector: M13, 87%
Sequencing vector: plasmid, 13%
Chemistry: Dye-primer ET; 87% of reads
Chemistry: Dye-terminator Big Dye; 13% of reads

Assembly program: Phrap; version 0.990319
Consensus quality: 2158/4 bases at least Q40
Consensus quality: 2194/2 bases at least Q30
Consensus quality: 2215/1 bases at least Q20
Insert size: 231000; agarose-fp
Insert coverage: 226623; sum-of-ctnigs
Quality coverage: 4.11 in Q20 bases; agarose-fp
Quality coverage: 4.19 in Q20 bases; sum-of-ctnigs

NOTE: This is a 'working draft' sequence. It currently consists of 20 ctnigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the ctnigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1447: contig of 1447 bp in length
1448 1547: gap of unknown length
1548 2609: contig of 1062 bp in length
2610 2709: gap of unknown length
2710 4301: contig of 1592 bp in length
4302 4401: gap of unknown length
4402 7988: contig of 3587 bp in length
7989 8088: gap of unknown length
8089 10498: contig of 2410 bp in length
10499 10598: gap of unknown length
10599 13494: contig of 2896 bp in length
13495 13594: gap of unknown length
13595 17577: contig of 3983 bp in length
17578 17677: gap of unknown length
17678 20847: contig of 3170 bp in length
20848 20947: gap of unknown length
20948 24832: contig of 3885 bp in length
24833 24932: gap of unknown length
24933 29810: contig of 4878 bp in length
29811 29910: gap of unknown length
29911 37323: contig of 7413 bp in length
37324 37423: gap of unknown length
37424 47954: contig of 10531 bp in length
47955 48054: gap of unknown length
48055 60562: contig of 12508 bp in length
60563 60662: gap of unknown length
60663 76734: contig of 16072 bp in length
76735 76834: gap of unknown length
76835 91907: contig of 15073 bp in length
91908 92007: gap of unknown length
92009 111826: contig of 19819 bp in length
111827 111926: gap of unknown length
111927 132932: contig of 21006 bp in length
132933 133032: gap of unknown length
133033 154508: contig of 21476 bp in length
154509 154608: gap of unknown length
154609 185758: contig of 31150 bp in length
185759 185859: gap of unknown length
185860 227194: contig of 41336 bp in length.

FEATURES
SOURCE
1. 227194
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-900K7"
1. 1447
/note="assembly_name:Contig6"
1548. 2609
/note="assembly_name:Contig7"
2710. 4301
/note="assembly_name:Contig8"
4402. 7988
/note="assembly_name:Contig9"
8089. 10498
/note="assembly_name:Contig10"
10599. 13494

misc_feature /note="assembly_name:Contig11"
13595. 17577
/note="assembly_name:Contig12"
17678. 20847
/note="assembly_name:Contig13"
20948. 24832
/note="assembly_name:Contig14
clone_end:SP6
vector_side:right"
24933. 29810
/note="assembly_name:Contig15"
29911. 37323
/note="assembly_name:Contig16"
37424. 47954
/note="assembly_name:Contig17"
48055. 60562
/note="assembly_name:Contig18"
60663. 76734
/note="assembly_name:Contig19"
76835. 91907
/note="assembly_name:Contig20"
92008. 111826
/note="assembly_name:Contig21"
111927. 132932
/note="assembly_name:Contig22"
133033. 154508
/note="assembly_name:Contig23"
154609. 185758
/note="assembly_name:Contig24"
185859. 227194
/note="assembly_name:Contig25"
40078 a 38264 g 75058 t 1930 others

BASE COUNT 71864 a 40078 c 38264 g 75058 t 1930 others
ORIGIN

Query Match 0.7% Score 21: DB 2: Length 227194;
Best Local Similarity 100.0%; Pred. No. 21:
Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1146 ataccagaaacacacacaaa 1166
|||||
Db 182182 ATACGAAACAAACACAAA 182202

RESULT 48
AE003538
LOCUS
DEFINITION
AE003538 303367 bp DNA INV 06-Oct-2000
Drosophila melanogaster genomic scaffold 142000013386050 section 30
OF 54, complete sequence.
ACCESSION
AE003538 AE002602
VERSION
AE003538.2 GI:10727947
KEYWORDS
HTG.
SOURCE
ORGANISM
fruit fly
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 303367)
REFERENCE
AUTHORS
Adams,M.D., Celisner,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galie,R.F.,
Georgie,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
Brandon,R.C., Rogers,Y.H., Blazey,R.G., Champe,M., Pfeiffer,B.D.,
Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
Miklos,G.L., Abriil,J.F., Agbayani,A., An,H.J.,
Andrews-Plamkoc,C., Baldwin,D., Balley,R.M., Basu,A.,
Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borokova,D.,
Botchan,M.R., Bouck,J., Brokstein,P., Brothier,P., Burris,K.C.,
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
Rados,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,


```

mrna
/translation="MSQESGDRSPICVGLSESGOMPKNKETTMTTTLVSSSPQOKMTM
QMMNTLGH"
join(<183391..183408,183461..>183691)
/gene="CG14113"
/product="CT33709"
/db_xref="FLYBASE:Fpan0014113"
/db_xref="FLYBASE:Fpqn0040814"
/evidence=not_experimental
<183391..>183691
/gene="CG14113"
/map="70A-70A4"
/db_xref="FLYBASE:Fpan0014113"
/db_xref="FLYBASE:Fpqn0040814"
/evidence=not_experimental
join(183391..183408,183461..183691)
/gene="CG14113"
/note="CG14113 gene product"
/codon_start=1
/db_xref="FLYBASE:Fpan0014113"
/db_xref="FLYBASE:Fpqn0040814"
/evidence=not_experimental
/protein_id="AAFA9834.1"
/db_xref="GI:7294491"
SLSDQVMGIREPQNPVAVPSPFDMNDKKNKAVARDP"

Query Match      0.7%; Score 21; DB 3; Length 303367;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1666 aaatctgaagcataaaca 1686
|||||
Db 162175 AAATCTGAAGCATAAACA 162195

RESULT 49
AY027789/c      768 bp      mRNA      PRI      20-JUL-2001
DEFINITION      Homo sapiens CLANC (CLAN1) mRNA, complete cds.
ACCESSION      AY027789
VERSION      AY027789.1 GI:14324116
KEYWORDS
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 768)
Damiano,J.S., Stehlik,C., Pio,F., Godzik,A. and Reed,J.C.
Cian, a novel human ced-4-like gene
Genomics. 75 (1-3), 77-83 (2001)
JOURNAL
MEDLINE
21365712
PUBMED
11472070
REFERENCE
2 (bases 1 to 768)
Stehlik,C., Damiano,J.S., Pio,F., Godzik,A. and Reed,J.C.
Direct Submission
Submitted (21-FEB-2001) Program on Apoptosis and Cell Death
Research, The Burnham Institute, 10901 North Torrey Pines Road, La
Jolla, CA 92037, USA
JOURNAL
FEATURES
source
Location/Qualifiers
1..768
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2p22-p21"
/tissue_type="lung"
1..768
/gene="CLAN1"
277..747
/gene="CLAN1"
/codon_start=1
/product="CLANC"

```

```

BASE COUNT      218 a      157 c      180 g      213 t
ORIGIN
1
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

Query Match      0.6%; Score 20; DB 9; Length 865;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 740 gctgctgaagctgcgcaga 759
|||||
Db 800 GCTGCTGAAGCTGC GCAGCA 819

RESULT 51

```

AC047412 1092 bp DNA HTG 13-APR-2000
 LOCUS Giardia intestinalis clone K10936 strain WB-C6, LOW-PASS SEQUENCE
 DEFINITION SAMPLING.
 AC047412 GI:7548543
 VERSION HTG; HTGS_PHASE0.
 KEYWORDS Giardia intestinalis.
 SOURCE Giardia intestinalis.
 ORGANISM Eukaryota; Diplomonadida; Hexamitidae; Giardia; Giardia.
 REFERENCE 1 (bases 1 to 1092)
 AUTHORS Morrison,H.G., McArthur,A.G., Nixon,J., Eakin,N.Q., Kim,U., Crocker,M.C., Hinkle,G., Holder,M.E. and Sogin,M.L.
 TITLE Giardia: a model for ancient eukaryotic genome analysis
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1092)
 AUTHORS Kim,U., Morrison,H.G., McArthur,A.G., Nixon,J., Eakin,N.Q., Crocker,M.C., Hinkle,G., Holder,M.E. and Sogin,M.L.
 TITLE Direct Submission
 JOURNAL Submitted (13-APR-2000) Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, 7 MBL Street, Woods Hole, MA 02543-1015, USA
 COMMENT * NOTE: This record contains 1 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved. 1092: contig of 1092 bp in length.
 * 1
 * Location/Qualifiers
 * 1..1092
 * /organism="Giardia intestinalis"
 * /strain="WB-C6"
 * /db_xref="taxon:5741"
 * /clone="K10936"
 *
 BASE COUNT 274 a 270 c 277 g 268 t 3 others
 ORIGIN
 Query Match 0.6%: Score 20; DB 2; Length 1092;
 Best Local Similarity 100.0%; Fred. No. 85;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 740 gctgtgaagctgcgcaga 759
 ||||||||||||||||
 Db 123 GCTGCTGAGCTGCGCAGA 142
 RESULT 52
 LOCUS REOS3NSB/C
 DEFINITION Reovirus serotype 2 S3 segment nonstructural protein (sigma-NS)
 RNA, complete cds.
 ACCESSION M18390
 VERSION M18390.1 GI:333757
 KEYWORDS sigma-NS nonstructural protein.
 SOURCE Reovirus serotype 2 (strain D5/Jones) viral RNA, passed in mouse L fibroblast cells.
 ORGANISM Mammalian orthoreovirus 2
 Viruses; dsRNA viruses; Reoviridae; Orthoreovirus; Mammalian orthoreoviruses.
 REFERENCE 1 (bases 1 to 1198)
 AUTHORS Wiener,J.R. and Joklik,W.K.
 TITLE Comparison of the reovirus serotype 1,2, and 3 S3 genome segments encoding the nonstructural protein sigma-NS
 JOURNAL Virology 161, 332-339 (1987)
 MEDLINE 88072071
 FEATURES Location/Qualifiers

source 1..1198
 /organism="Mammalian orthoreovirus 2"
 /db_xref="taxon:10885"
 CDS 28..1128
 /note="nonstructural protein (sigma-NS)"
 /codon_start=1
 /protein_id="AAA47282.1"
 /db_xref="GI:333758"
 /translation="MASSIRRAISIKIKDDNGQGVNMYLRSSVNTKVRNVVDIQ
 IKTGFESCIAMLRPLQYAKRERLLGKRLERLARIADVLQTRDLHSLCMPEDAPMT
 YQASTMRLEIVCDHFVHDVGLVYPMDDRSPPSLARLFTGWAGLITTEPAKRV
 PIMHLAALDCCWTFALPYMITVVDGTVVAPVATPLAERLLDGGFGYCLDYSYCEV
 DANNASAGDSMSDSRCINETVTAETAEACILKICLILNCQKFLKMDLADHNFEE
 DKVQMTPEFSEFVPMASAPNATVDVOCFRCILKMDKMLKIDMRMTMLWRAGSDDA
 ISTSLTISLDGRVYADMDNEVRLVTPARY"
 BASE COUNT 288 a 272 c 306 g 332 t
 ORIGIN
 Query Match 0.6%: Score 20; DB 14; Length 1198;
 Best Local Similarity 100.0%; Fred. No. 85;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2179 attctctcatgttggaagcc 2198
 ||||||||||||||||
 Db 358 ATTCTCTCATGCTGAGAACCC 339
 RESULT 53
 LOCUS SCYBR141C/C
 DEFINITION S. cerevisiae chromosome II reading frame ORF YBR141C.
 ACCESSION Z36010.1 Y13134
 VERSION Z36010.1 GI:536439
 KEYWORDS
 SOURCE baker's yeast.
 ORGANISM Saccharomyces cerevisiae
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 REFERENCE 1 (bases 1 to 1688)
 AUTHORS Becam,A.M., Herbert,C.J., Nasr,F., Slonimski,P.P. and Zagulski,M.
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1688)
 AUTHORS MIPS.
 TITLE Direct Submission
 JOURNAL Submitted (30-AUG-1994) Data collected by MIPS on behalf of the European yeast chromosome II sequencing project. MIPS at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152 Martinsried, FRG; E-mail: Mewes@mips.emblnet.org
 3 (bases 1 to 1688)
 Feldmann,H., Aigle,M., Aljinovic,G., Ande,B., Bactlet,M.C., Barthe,C., Baur,A., Becam,A.M., Bileau,N., Boles,E., Brandt,T., Brendel,M., Bruckner,M., Bussereau,F., Christiansen,C., Contreras,R., Crouzet,M., Cziepluch,C., Demolis,N., Delaveau,T., Dolignon,F., Dondy,H., Duesterhus,S., Dubois,E., Dujon,B., El Bakoury,M., Entian,K.D., Feuerhahn,M., Fiers,W., Fobg,G.M., Fritz,C., Gassenhuber,H., Giansdorff,N., Goffeau,A., Grivell,L.A., de Haan,M., Hein,C., Herbert,C.J., Joniaux,J.C., Holmstrom,K., Jacq,C., Jacquet,M., Jauniaux,J.C., Honiaux,J.L., Kallioe,T., Klesau,P., Kirchath,L., Koelter,P., Korol,S., Liebl,S., Logghe,M., Lohan,A.J.E., Louis,E.J., Li,Z.Y., Mat,M.J., Mallet,L., Mannheim,G., Messenguy,F., Miosga,T., Molemans,F., Mueller,S., Nasr,F., Obermaier,B., Perau,T., Plerard,A., Piravandi,E., Pohl,F.M., Pohl,T.M., Potier,S., Proft,M., Purrelli,B., Ramezani Rad,M., Rieger,M., Rose,M., Schaefli-Geertschlaeger,I., Scherens,B., Schwarze,C., Skala,J., Slonimski,P.P., Smits,P.H.M., Souciet,J.L., Steensma,H.Y., Stucka,R., Urestarazu,A., van der Aart,O.J., van Dyck,L., Vassarotti,A., Vetter,I., Vierendeels,F., Vissers,S., Wagner,G., de Wergifosse,P., Wolte,K.H., Zagulski,M., Zimmermann,F.K., Mewes,H.W. and Kleine,K.
 TITLE Complete DNA sequence of yeast chromosome II
 JOURNAL EMBO J. 13 (24), 5795-5809 (1994)
 MEDLINE 95112788
 FEATURES

REFERENCE 2 (bases 1 to 3709)
AUTHORS Bourbon,H.-M.
TITLE Direct Submission
JOURNAL Submitted (23-DEC-1998) UMR5540 du CNRS, CNRS, 118 Route de
Narbonne, Toulouse 31062, France
FEATURES
source
1. 3709
/organism="Drosophila melanogaster"
/strain="Oregon R"
/db_xref="taxon:7227"
/chromosome="2"
/map="2R: 448"
join(257..562,761..972,1228..1629,1780..1913,1979..2046,
2472..3226)
/gene="Pabp2"
/product="poly(A)-binding protein II"
257..3226
/gene="Pabp2"
/gene="Pabp2"
join(874..972,1228..1629,1780..1913,1979..2018)
/gene="Pabp2"
/function="RNA-binding protein"
/codon_start=1
/product="poly(A)-binding protein II"
/protein_id="AAE00976.1"
/db_xref="GI:6007612"
/translation="MADEDITINEDDLEETNGEOTETATEVEEGSMOIDEPL
EATKARKMEEEAEKIKOMOSEVDKMGSGTGLATVPLSLEKOEIDTSSVYVGN
VDYGASAELEAHFHCCGTTNNYITLCKNADGPKGFAITEGSKFEVFTALAMNETL
FGKROIKVSKRITNRPGLSTTNFAPGSRFGARVSRACHSTFGARRAMGYRGA
NYAPY"
BASE COUNT 1146 a 777 c 785 g 1001 t
ORIGIN
Query Match 0.6%; Score 20; DB 3; Length 3709;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2534 tgcgaatgcagtgaaatcc 2553
|||||
Db 684 TGCNAATGCAGTGAATAATCC 665
RESULT 56
SCIRAI/c
LOCUS SCIRAI 12595 bp DNA PLN 22-SEP-1994
DEFINITION S.cerevisiae (S288C) IRA1, YBR1118 and YBR1119 genes.
ACCESSION X78937
VERSION X78937.1 GI:547575
KEYWORDS DEAD box protein; helicase; IRA1 gene; YBR1118 gene; YBR1119 gene.
SOURCE baker's yeast.
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE 1 (bases 1 to 12595)
AUTHORS Zagulski,M., Becan,A.M., Grzybowska,E., Lacroite,F., Migdalski,A.,
Slonimski,P.P., Sokolowska,B. and Herbert,C.J.
TITLE The sequence of 12.5 kb from the right arm of chromosome II
predicts a new N-terminal sequence for the IRA1 protein and reveals
two new genes, one of which is a DEAD-box helicase
JOURNAL yeast 10 (9), 1227-1234 (1994)
MEDLINE 95274325
REFERENCE 2 (bases 1 to 12595)
AUTHORS Herbert,C.J.
TITLE Direct Submission
JOURNAL Submitted (21-APR-1994) C.J. Herbert, Centre de Genetique
Moleculaire, CNRS, 91198 Gif-sur-Yvette, FRANCE
FEATURES
source
1. 12595
/organism="Saccharomyces cerevisiae"
/strain="S288C"
/db_xref="taxon:4932"
/chromosome="II"

gene
complement(1..8331)
/gene="IRA1"
CDS
/gene="IRA1"
/codon_start=1
/protein_id="CAA55537.1"
/db_xref="GI:547576"
/translation="MNQSDPOKKNPMPMYSTKHLFEDRLIVLPISNKTAYDVE
ADSVNCSSTILNTAITKDLNPITENTIGLIDLVODEBITSDNTIDTIAISIVLV
RLSDVFEYVWQNDNDEKIRNDONRPGKSPNPFHPSRPHATIRNPLATMLLCKI
SKIRKFTRLKVLQNNSHLSCSARISKSIIPDSEOFQOKNPNAYAEIKEDLTIDYI
OREISANIVEFTKCVKTVVAPLILSHSTFGEVYVNHLDLFCGECNLTALADI
LOHLSYMKRTTFEHLILYYASKAFPLIMAPKRYKTYNNLSIDSDNPPSSDNG
GSNNKSTISOLVSLIPDDVYSTFSVSLTNVNDHHYHLHSSSSSKTTNTNSPN
STSKTISKOSVYASGNVSPSGTGNDSPPSPAPASLSPPLNTNITLGLPSITSTL
GOANTSTSTTAATRTDADPTPTMTNTNNNNNSANLNINIPORLFSLDDISSFNSR
KSLNLDSSNLSFLMDTSHSNASMTNTNNHAGVNNNSQSDSSNLMENIMELYSN
TGSLSHSTALIRPLVLTLLDSEYDEKNSYKIRSEPININPKDSNTSMGAS
KNPSIRHLTHGLKTLIOGGRNRYKFLTYLRNNGQOFVDSVLDSIRSLFLMT
MPSISQIDSNASVIFSKRFLNLLGONLEVNTNMGATANTFISHCYBRNPLTHRL
OLEFPAQGLQDLSLFLRHLQLEKLNHLDLPLISLYTEGEFVFFHLVSTKLHSDIA
EKTSVLRKLFPCIIADILKATPYDDVNTKLIASLDGHIIDQEDARTSLNDHVS
PDATSVYEPTEPIIHNSSDASLVSLSSPLSINGSGITWTFTWIDIOSILPILSNR
SSASDLSNLTINPLEAQQNNANILARLSGVPTRKRYASPNSERSROSPSSP
OLOOSDLSPLSLVSSAGFSSNHSTTATPTLTKIKSPKPKTKKIDADQOLQPSY
SVYTLSDNDEAKIMANITSTIKRNTNFIIRDANTEPKTTDILKPLFVSLDSNQ
RLOVAPARAFETLPLTAFEDIDNDLDPRLVDHLYLCTYAVTLFASLPLKLENA
KREMLDIIIVKFORVRSYLSNLAEKHNLOVALITTEIRLPLLVAGVSGILFSLYCS
RNTPRRLIKISCEEFRLRFQKYVGLDOYIYNIDFIDMAQDNFASGVVALOR
RLRNNTLITYIKGSDSLDSMDVIYKKMFYFCSKSVQOEVLDPRLSLAGILASGSI
LSDMOLEKRSKAPNEDGSLSFESRNPAVEYKSLKLELTCKMNFISKOCOMITP
NLITRENSRDIITSLHPLSLFNNIGLKTIDELMSIDLSKSHDSSSVLLEQIITI
ITRTIKRDDDEKIMLFSTDLDAVDKLEIYEKISTSKYKGTIQNSKFRPEH
SEKNIGLISNHFHLKMKLVLGVGRKLSINKQYDEENSRPLREMDLQRDPLFYID
TSEISAKALVLTNHPLEIPSSSEKEDNRSTVSFGNHFITLKGLEKSDMLQFP
VSLRHSITLNNVITILNLSNANVNSLKTLPNGYSPNDIRIAFLRYVIDVTA
YVNPKEHEMDKMLAIDDLKYITKNPILAFGSLGAPDADVAGGLNFAFDRNA
SHIIVTEILKOEIKRARSDDLRRNSCATRALSTYRGRGKKYILKTLRPVLOGID
NKESEFIDKMGKGSENSEKMDLPEKRYRLIDATYSSIDDPETELVDICTITYAAS
VNPPEYAVIAGSVFVFLRIFGPAVSPDSENIIVYHADRPFTTLAVIOLSLANGR
ENIFKDDILVSKKEELKTCSDKIFNPLSELCKIPITNPFVNSRFTSIRKFNVYKIS
FYLNEFTIRKEIINESKLPGSEFLKNTVMNDKILGVQPSMEIKMEIIPFVEN
REKYSIYEFMSRYAFKVKYDMEEBEDNAPEHAWMTLDGIIQIIVTFPNCYNNV
DSLYVKVIOIYARMCSKRYVIDCTTPPGKANGKLTLEFSLIPBASNCAGCY
YFNVNSFMDQWASSTYENPLVLTTPPCFINSNTDOSLISLGSLSGLSLVDOR
VTLHDITLIDKERRKFCVSLKIGNKYEQVLEHIDQLYKVTYSNRTFSIKFNVYKIS
NLISVDVNTTGVSSSEFTLIDNEKLYGSEFKYLEIVKMFYQAQKMEDEGDTFSN
DISFSTSSAVNASCYCNVKEGEIITSHSLVILVGLFENDDLKVNINSYLLAVAGDAF
NIDFGTRLKSPETVYPDOTTFALILFKAFSSSESTELPPIWKMLOGLENDVPOE
HPTVYCSLSYVNPNTYEHVYLANDDEGEALSRITISLRTYKPNPTTYAQOIN
FLIADGRITNVIETIVSHALDROSENDDMKAAVILSPFTEIACOVILKIMTI
KSFPLAVEASAHWSLTLTISKISVSLFESSPLLSQWYDEILFVAVSLIDVDPSE
IIVSYLIELMNVCHLTLNNESLPERNRKMLDLYCAOKLNFISQFQEGEYAPLN
FAASSFSKFEGLIDFTKNIMLMEYGSISEAOEAKYKTYLMDAIFGRHFFSARA
NMIGIMSKSHSLFLCKELLYETKVFPEVYVDDQMMIIIAHFTYSKIYEGDPS
SELMKEL"
complement(8728..9741)
/gene="YBR1118"
CDS
complement(8728..9741)
/gene="YBR1118"
/codon_start=1
/protein_id="CAA55538.1"
/db_xref="GI:547577"
/translation="MHSRKSITGRKRVGNSVTVIRVIRPQXTRIRIRRHILINKRQ
SICKFLIKENDDSNEEKNDKIRISIGANRYEYEDGSGOSFNDMEQLRLHL
SLIKNSKSKOTSDLAIVYTLIGYIMNOIGLGEYETIOIASONQOLKRGDTSKLL
EKWIRSSPENCAGVALEITGSSGRIRISRCALFRVNYVIDLEEHGKVOQPMRPL
PRNENDFDLISCLVLANFVKNRHROGACHMVRFLAKOGITIFVLPOACTYHSYC
DTLLQNLGSLIGLIMLNSHNSKNLYCYLOLVVSPDSSFSKRIKVDGGLNNG
ITL"

250110.
The end of this sequence (40834..40937) overlaps with the start of
sequence 270208.
For a graphical representation of this sequence and its analysis
see: - [http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?](http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=R01E6)
name=R01E6.

FEATURES

SOURCE

Location/Qualifiers

1..40937

/organism="Caenorhabditis elegans"

/db_xref="taxon:6239"

/chromosome="X"

/clone="R01E6"

join(999..1127,1467..1589,1647..1756,1803..1905,
1958..2060,2118..2307,2362..2503,2557..2747,3149..3251,
3304..3441,3489..3656,4047..4244)

/gene="R01E6.4"

join(999..1127,1467..1589,1647..1756,1803..1905,
1958..2060,2118..2307,2362..2503,2557..2747,3149..3251,
3304..3441,3489..3656,4047..4244)

/gene="R01E6.4"

/note="contains similarity to Pfam domain: PF00065
(Neurotransmitter-gated ion-channel), Score=484.8,
E-value=2.2e-142, N=1

/protein_id="CA92184.1"

/db_xref="GI:3878769"

/translation="MSEFWMALILLVAVGTNAKIRKTAQLESQYEDLLFDYNY
PRPVNSSDILTVVGASLRIIDVDEKNQVLTNNLEMKMNDAKLWPEKYGKLT
TLHPSPDITPDLVLYNNAAGDDITLTDALVTBGNVWOPAIKSPCIDVTW
EPYDSCCKEMFGWTYGRVYDKQLEQEEVITIKDNDVEFMQOQMDSPFYRSA
EMDLTLSEHRSVLYASCQPEKRYVDITTYEGRKTLFTCNILILCPILSTLT
VFYLSDKITPESISITVLYFLVLDLMDPTSLVPMERGLITTMILVALSTVS
VITYNFRSSGAHAKSBWIAVFLKFLKLLMSRPEKGEVTKOPLVNASTLGSN
YATAKANERYRNKAKGNNDLMSLRGPSSQAVLNFTEDMRNRNNDSTVNRKPYL
CAQNNAPVAFERKTRKRSKSDVDVFMNLNQVRFJAEHFRNELGEISDMTETV
AMVLDRLFLIFLVNLTGVTIFLLESPSLYDSKPMNITVPNKLQGNMFSSNMK"

join(5029..5054,5131..5546,5594..5889)

/gene="R01E6.5"

join(5029..5054,5131..5546,5594..5889)

/gene="R01E6.5"

/note="CDNA EST YK146b3.5 comes from this gene"

CDNA EST YK170c5.5 comes from this gene

CDNA EST YK146b3.3 comes from this gene

CDNA EST YK170c5.3 comes from this gene

CDNA EST YK412b8.3 comes from this gene

CDNA EST YK412b8.5 comes from this gene

CDNA EST YK387b5.3 comes from this gene

CDNA EST YK422c1.3 comes from this gene

CDNA EST YK422c1.5 comes from this gene

CDNA EST YK309d12.3 comes from this gene

CDNA EST YK309d12.5 comes from this gene

/codon_start=1

/protein_id="CA92185.1"

/db_xref="GI:3878770"

/translation="MASPISHSHENASGKYGDESHCKNPHOKAHSHEKAKG
KYSQGRKADHDALHKADEHKONKEAHQSGSKNYVDASNTNDVKTFTGFEDRY
VOPYHMEQYHTDERHANKYAGDEHNAGQKRDGYNADHOGYDKAQEHGASHTN
YGNEDAGHSKYDNECKYGDHDTGYNKGYNKGYDNGCGSDSESYQPSYQPKYGGH
QSHYQPSARHSSPOGDOYGYEAPH"

join(11014..11029,11105..11235,11321..11420,11472..11686,
11940..12167,12215..12304)

/gene="R01E6.3"

join(11014..11029,11105..11235,11321..11420,11472..11686,
11940..12167,12215..12304)

/gene="R01E6.3"

/note="contains similarity to Pfam domain: PF00194
(Eukaryotic-type carbonic anhydrase), Score=144.1,

E-value=8e-40, N=1
CDNA EST YK202f1.3 comes from this gene
CDNA EST YK202f1.5 comes from this gene
CDNA EST YK234h4.3 comes from this gene
CDNA EST YK234h4.5 comes from this gene

/codon_start=1

/protein_id="CA92190.1"

/db_xref="GI:3878775"

/translation="MPRRKREKSKTSSRRRLKTLTPFLAAROSPIDIPOHVCCD
TVCQKALANTDYKSGCCDYLVSDEGGLVAVKRNCKEFTLANHLPSSKALQPAFH
WGSNKSRESSEHFLDGKLSGVEHVEFNNTSLPESKVALSKDGLAVYGVFLKEKTYD
NYHGLIDVFRKATGNAPFIAMKDFHIEHLPSDKREVEYVLSLTPPVNECVIYT
LPEPEVSEFQGLVLRNIIIPANRAQCDRCDEIRRSFN"

complement(join(13219..13506,13552..13698,13747..13815,
13862..13985,14486..14626,14675..14843,14895..15086,
15323..15497,15549..15620,15673..15963,16018..16099,
16292..16364,16408..16501,16800..16951,17049..17172,
17430..17501,17549..17721,17789..17885,17932..18006,
18054..18420,18589..18656))

/gene="R01E6.1"

complement(join(13219..13506,13552..13698,13747..13815,
13862..13985,14486..14626,14675..14843,14895..15086,
15323..15497,15549..15620,15673..15963,16018..16099,
16292..16364,16408..16501,16800..16951,17049..17172,
17430..17501,17549..17721,17789..17885,17932..18006,
18054..18420,18589..18656))

/note="contains similarity to Pfam domain: PF00069
(Eukaryotic protein kinase domain), Score=25.7,
E-value=1.0e-06, N=2; PF00211 (Adenylate and Guanylate
cyclase catalytic domain), Score=191.2, E-value=5.5e-54,
N=1"

/codon_start=1

/protein_id="CA92186.2"

/db_xref="GI:14530534"

/db_xref="SPRMBL:Q21617"

/translation="MCVRLQWRMPDSVOOQSOYSEKHLELLTVILKLFGEYHRI
NQOQSGSDVNSKASAYAINAVARSTSGELDFVYGPCTDITTDIDFEIMKSPYI
GEYVFEARQVQELIYVNAQFSGVAETVFLMELEQVETTLGSKYKLVNGIS
LSNDLRSENEIMSEKFIREYVEVDENVDMTRKQKRGAMVAVOCADYDIASVY
NIGTRLSLGFPEITVILNKPPDEILNOPNKLTVSNMFLTSPLOEQSDAISOQ
DVIPNLADDOGTFITLRIHACIACVSVNKALETQTDNIHTNASKAVYTTGCTTDD
NSGSVLTVNAVETINPAMTFESLITLTVSKASCDYTNCSQLSNKRKSDILMTKTDMD
PPDDCAKSSCVNYIPIHIAVAVIYIIVAVIIVAKRRLIYKLTWKPKRESIK
IIVKNKNDKQRELENRASNTDANALSRFVGSYALVGTQRAVYQFOKRIKN
FPEITLVLYLSKLOLDNLAKFYGIQVNDIMPTLHTLTVGCTLEECDDDFCM
DDTFKSAFMRDITLKGLOYLHKRSIGYGHIOASTCLDIDIMVYKLTIGVSNFMSDL
DAENIKYPEQAHMTTYPOYCPPEHIREYDSDGCKPPRVRSRPGDILYCGMITY
MVRREDPYHLIHSVERPNATLKOIILNENMPRTTDDYRQENMLLMCKECKWRND
KRPTIKRLIESISTVYPLSKGNLVDOMIRMSKEXYADELQWVALRTADLADAOQWTR
LLNMLPASIAKDKLNGLIIMPRESYATVAFVQICDFNALMKRSSPEOVYAFINDY
DOFTVIRKRDYAVYETTGETVAVASGVPHENEGRHLEFEVAEISLETRETSYIVLDH
DKNYKLIRIGFHAGPIAAGYIGIRSPRYCIGGTVAFSMOSNCPNPDQISLITA
RLLDSEHKYKVGKGIYVAVKGNAAALIKCCEFFEHSHDL"

complement(join(20486..20551,21018..21101,21336..21491,
21729..21881,22959..23081,23639..23753,23802..23950,
23999..24088))

/gene="R01E6.6"

complement(join(20486..20551,21018..21101,21336..21491,
21729..21881,22959..23081,23639..23753,23802..23950,
23999..24088))

/gene="R01E6.6"

/note="similar to 951004: R01E6.6 has homology with
C368.2"

/codon_start=1

/protein_id="CA92187.1"

/db_xref="GI:3878772"

/db_xref="SPRMBL:Q21618"

/translation="MGNSSSSSRINKKSEMEIRKPNRSSVSGFEFNKSKSMMDSS
KSSMSTASNTSRKSTISIKKDKPKRSKNEISRTSKMRDSDPISAGGEE
ITTCGFENPHSEFRANKYVQRIFEKREDYQKTIIMLUGERSIYVNNRKQVLEIVAH
HDADFLSVSKYGEHEHVKQYGFDPDVAVADAMTLEGVILDMANNHAPDVSAM

gene

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

gene

SLVMTIFFSVADGYVSELRRHRMSSRLTKHSTVDSENNEDTPEIHSVCLHI
PLSPHIIIPANEVYTLQRLLOQAAPARSSNNL"
Complement(join(33511..33756,33804..33947,33995..34096,
34158..34242,34362..34369))
/gene="R01E6.2"
34158..34242,34362..34369))
/gene="R01E6.2"
/codon_start=1
/protein_id="CAA92188.1"
/db_xref="GI:3878773"
/db_xref="SPTREMBL:Q21619"

CDS

Query Match 0.6%; Score 20; DB 3; Length 40937;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2414 gttgtattcattgacc 2433
Db 22674 GTGTTATTTCATTGACCC 22693

RESULT 59

LOCUS CBRG39N14 45027 bp DNA INV 04-NOV-2000
DEFINITION Caenorhabditis briggsae cosmid G39N14, complete sequence.
ACCESSION AC084575
VERSION AC084575.1 GI:11095025
KEYWORDS HTG.
SOURCE Caenorhabditis briggsae.
ORGANISM Caenorhabditis briggsae.
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 45027) Washington University Genome Sequencing Center.
The C. briggsae genome sequencing project
Unpublished
2 (bases 1 to 45027)
Waterston, R.
Direct Submission
Submitted (04-NOV-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
JOURNAL
TITLE
AUTHORS
REFERENCE
COMMENT
Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA
e-mail: jsplth@watson.wustl.edu

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA
e-mail: jsplth@watson.wustl.edu

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.

FEATURES

Location/Qualifiers
1..45027
/organism="Caenorhabditis briggsae"
/strain="Gujarat 616"
/db_xref="taxon:6238"
/clone="G39N14"

BASE COUNT 14506 a 8254 c 8601 g 13666 t
ORIGIN

Query Match 0.6%; Score 20; DB 3; Length 45027;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 23 ttatccagaacaatgaatt 42
Db 2861 TTTCATCAGACATGATT 2842

RESULT 60

AC006103

LOCUS AC006103 45459 bp DNA HTG 04-DEC-1998
DEFINITION Homo sapiens chromosome 10 clone LA10NC01_124_D_3 map 10q25.1, ***
SEQUENCING IN PROGRESS ***, 1 ordered pieces.
ACCESSION AC006103
VERSION AC006103.1 GI:3962484
KEYWORDS HTG; HTGS_PHASE2.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 45459)
Smith, D.R.
Sequencing of Human Chromosome 10
Unpublished
2 (bases 1 to 45459)
Smith, D.R.
Direct Submission
Submitted (04-DEC-1998) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02154, USA
Note: Clone was sequenced in 1996 using Multiplex DNA Sequencing
technology. Data may contain low quality seq unce and BAC/Cosmid
vector sequences.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 45459: contig of 45459 bp in length.
Location/Qualifiers
1..45459
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="LA10NC01_124_D_3"
/chromosome="10"
/map="10q25.1"

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Submitted (04-DEC-1998) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02154, USA
Note: Clone was sequenced in 1996 using Multiplex DNA Sequencing
technology. Data may contain low quality seq unce and BAC/Cosmid
vector sequences.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 45459: contig of 45459 bp in length.
Location/Qualifiers
1..45459
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="LA10NC01_124_D_3"
/chromosome="10"
/map="10q25.1"

FEATURES

Location/Qualifiers
1..45459
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="LA10NC01_124_D_3"
/chromosome="10"
/map="10q25.1"

BASE COUNT 13350 a 9755 c 9636 g 12671 t 47 others
ORIGIN

Query Match 0.6%; Score 20; DB 2; Length 45459;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2914 tattgagaatcttaagca 2933
Db 8207 TATTGAGATCTTAAGCAA 8226

RESULT 61

LOCUS AC005210 83969 bp DNA PRI 05-NOV-1999
DEFINITION cith 179_n_3, complete sequence.
ACCESSION AC005210
VERSION AC005210.3 GI:6249673
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 83969)
Smith, D.R.
Sequencing of Human Chromosome 10
Unpublished
2 (bases 1 to 83969)
Smith, D.R.
Direct Submission
Submitted (01-JUL-1998) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02154, USA

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Submitted (01-JUL-1998) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02154, USA

REFERENCE 3 (bases 1 to 83969)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (11-DEC-1998) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA

REFERENCE 4 (bases 1 to 83969)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (02-MAR-1999) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA

REFERENCE 5 (bases 1 to 83969)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (05-NOV-1999) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA

REMARK COMMENT Vector Sequence Clipped
FEATURES On Nov 5, 1999 this sequence version replaced gi:4314329.
location/Qualifiers
1..83969
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone_id="RPCI-11.2"
BASE COUNT 24380 a 19036 c 18255 g 22298 t
ORIGIN

Query Match 0.6%; Score 20; DB 9; Length 83969;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2914 tatttgagaatcctaagcaa 2933
|||||
Db 68688 TATTGGAATCTTAAGCAA 68707

RESULT 62
AL159169/c 86155 bp DNA PRI 28-AUG-2000
LOCUS Human DNA sequence from clone RP11-408A13 on chromosome 9, complete sequence.
DEFINITION
ACCESSION AL159169
KEYWORDS HMG.
SOURCE
ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 86155)
AUTHORS Collier,R.
TITLE Direct Submission
JOURNAL Submitted (28-AUG-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
COMMENT On Aug 29, 2000 this sequence version replaced gi:9908927.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; SW: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human

Chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr9>
RP11-408A13 is from the library RPCI-11.2 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/>
VECTOR: pBAC3.6
IMPORTANT: This sequence is not the entire insert of clone RP11-408A13. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.
The true right end of clone RP11-408A13 is at 86155 in this sequence. The true left end of clone RP11-439M12 is at 95 in this sequence. The true right end of clone RP11-120J1 is at 100 in this sequence.

FEATURES
source
1..86155
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-408A13"
/clone_id="RPCI-11.2"
complement(1..92)
/note="match: GSS: Em:AQ351225"
104..499
/note="match: GSS: Em:AQ830380"
113..440
/note="match: GSS: Em:AQ720173"
620..831
/note="106 copies 2 mer tt 55% conserved"
921..1076
/note="MIR repeat: matches 2..168 of consensus"
1112..1245
/note="L1M4 repeat: matches 4699..4836 of consensus"
1511..2117
/note="match: GSS: Em:AQ427218"
2745..3053
/note="AluSg repeat: matches 1..309 of consensus"
3152..3341
/note="MIR repeat: matches 48..247 of consensus"
3697..3752
/note="28 copies 2 mer ac 85% conserved"
4146..4327
/note="MER5A repeat: matches 7..178 of consensus"
4708..4847
/note="MIR repeat: matches 92..241 of consensus"
4899..5065
/note="MER5A repeat: matches 3..189 of consensus"
5103..5207
/note="MER5A repeat: matches 65..166 of consensus"
5132..5233
/note="MER5A repeat: matches 9..112 of consensus"
7190..7282
/note="L2 repeat: matches 2331..2421 of consensus"
7384..7439
/note="L2 repeat: matches 2577..2634 of consensus"
9060..9293
/note="MIR repeat: matches 3..243 of consensus"
9777..10318
/note="match: GSS: Em:AQ395702"
complement(10404..10831)
/note="match: GSS: Em:AQ030903"
10804..10990
/note="MIR repeat: matches 47..231 of consensus"
14497..14804
/note="match: GSS: Em:AQ665295"
14510..15126
/note="match: GSS: Em:AQ310218"
14782..14976
/note="MIR repeat: matches 66..256 of consensus"
15854..16584
/note="match: GSS: Em:AQ394111"
17012..17077
/note="L2 repeat: matches 2680..2745 of consensus"

```

repeat_region 19799..19828
/note="15 copies 2 mer ga 93% conserved"
repeat_region 20263..20342
/note="40 copies 2 mer aa 68% conserved"
repeat_region 23323..23412
/note="12 repeat: matches 2594..2688 of consensus"
repeat_region 24826..24898
/note="MIR repeat: matches 46..119 of consensus"
repeat_region 24991..25058
/note="34 copies 2 mer ga 67% conserved"
repeat_region 25396..25600
/note="MIR repeat: matches 1..217 of consensus"
repeat_region 26063..26184
/note="12 repeat: matches 2572..2710 of consensus"
repeat_region 27687..27775
/note="LIM3 repeat: matches 7659..7739 of consensus"
repeat_region 28585..28788
/note="12 repeat: matches 2558..2749 of consensus"
repeat_region 28834..29034
/note="AluY repeat: matches 1..201 of consensus"
repeat_region 29035..30250
/note="LIM4 repeat: matches 3018..4252 of consensus"
misc_feature 29792..30537
/note="match: GSS: Em:AQ353410"
misc_feature 29795..30640
/note="match: GSS: Em:AQ898185"
misc_feature 30938..31350
/note="match: GSS: Em:AQ627562"
repeat_region 32743..32817
/note="MIR repeat: matches 187..262 of consensus"
repeat_region 33146..33212
/note="12 repeat: matches 2679..2750 of consensus"
repeat_region 35009..35139
/note="LIM5A repeat: matches 6162..6294 of consensus"
repeat_region 35486..35885
/note="MSTB repeat: matches 1..426 of consensus"
repeat_region 35886..37460
/note="MSTB-internal repeat: matches 1..1651 of consensus"
repeat_region 37463..37675
/note="MSTB repeat: matches 203..420 of consensus"
repeat_region 37685..37913
/note="MSTB repeat: matches 1..229 of consensus"
repeat_region 39397..39444
/note="24 copies 2 mer tt 75% conserved"
repeat_region 40066..40259
/note="LIM4 repeat: matches 4385..4593 of consensus"
repeat_region 40534..40833
/note="LIM3 repeat: matches 5866..6180 of consensus"
repeat_region 41355..41394
/note="20 copies 2 mer aa 82% conserved"
repeat_region 42168..42231
/note="12 repeat: matches 2642..2705 of consensus"
repeat_region 43215..43505
/note="AluY repeat: matches 1..307 of consensus"
repeat_region 43933..43969
/note="12 repeat: matches 2696..2732 of consensus"
repeat_region 43935..43986
/note="MIR repeat: matches 216..260 of consensus"
repeat_region 44879..45241
/note="12 repeat: matches 2264..2709 of consensus"
repeat_region 45313..45472
/note="MIR repeat: matches 4..183 of consensus"
repeat_region 45633..45946
/note="AluY repeat: matches 3..312 of consensus"
repeat_region 46261..46675
/note="HAI1 repeat: matches 1098..1552 of consensus"
repeat_region 46676..46713
/note="19 copies 2 mer aa 81% conserved"
repeat_region 46993..47895
/note="12 repeat: matches 1181..2181 of consensus"
repeat_region 48487..48532
/note="23 copies 2 mer ta 76% conserved"
repeat_region 48581..48646

```

```

repeat_region /note="33 copies 2 mer at 68% conserved"
48824..49187
/note="THE1C repeat: matches 1..371 of consensus"
repeat_region 49538..49616
/note="MERS8B repeat: matches 275..336 of consensus"
repeat_region 49617..49972
/note="LIM2 repeat: matches 5810..6163 of consensus"
misc_feature complement(50122..50460)
/note="match: GSS: Em:B88855"
repeat_region 50230..50439
/note="MER30 repeat: matches 5..230 of consensus"
misc_feature 50459..50934
/note="match: GSS: Em:AQ247716"
repeat_region 51761..52062
/note="MER33 repeat: matches 3..323 of consensus"
repeat_region 54014..54039
/note="13 copies 2 mer ca 100% conserved"
repeat_region 54903..54981
/note="MIR repeat: matches 51..139 of consensus"

Query Match 0.6%: Score 20; DB 9; Length 8615;
Best Local Similarity 100.0%: Pred. No. 76;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1962 tacattccagcagggctgt 1981
|||||
Db 62606 TACATTCCAGCAGGCTGT 62587

RESULT 63
AP000885 86719 bp DNA PRI 16-DEC-1999
LOCUS Homo sapiens genomic DNA, chromosome 21q22.1, clone:B680H4, SOD-AML
DEFINITION region, complete sequence.
ACCESSION AP000885
VERSION AP000885.1 GI:6580114
KEYWORDS HTG.
SOURCE Homo sapiens DNA, clone:B680H4.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 86719)
REFERENCE
AUTHORS Fujiyama,A., Yada,T., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Hattori,M., Ishii,K., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 86,719 genomic DNA of 21q22.1
Published Only in Database (1999) in press
2 (bases 1 to 86719)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (15-DEC-1999) to the DDBJ/EMBL/Genbank databases.
Masahira Hattori, The Institute of Physical and Chemical Research
(KIKEN), Genomic Sciences Center (GSC), Kitasato Univ., 1-15-1
Kitasato, Sagamihara, Kanagawa 228-8555, Japan
(E-mail:hattori@gs.c.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-42-778-9923, Fax:81-42-778-9924)

FEATURES
Location/Qualifiers
Source 1..86719
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/clone="B680H4"
/map="21q22.1"

BASE COUNT 26401 a 16034 c 16025 g 28259 t
ORIGIN
Query Match 0.6%: Score 20; DB 9; Length 86719;
Best Local Similarity 100.0%: Pred. No. 76;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2465 catatccagtcctgtcaa 2484
|||||

```

Db 3109 CATAGTCAGTCTCTGTCAA 3128

RESULT 64
AC005923
LOCUS
DEFINITION Homo sapiens 3p21.3-4 PAC RP4-751E10 (Roswell Park Cancer Institute
AC005923 complete sequence.
AC005923
VERSION AC005923.2 GI:4309927
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 88326)

REFERENCE
AUTHORS Muzny,D., Arenson,A.D., Bouck,J., Bunag,C., Chen,J., Chen,Z.,
Culpepper,P., Ding,Y., Dugan,S.P., Durbin,K.J., Forcum,J.,
Ganesh,R.P., Garcia,C., Garcia,D.K., Gorrell,H., Gorrell,L.L.,
He,X., Hernandez,J., Jackson,L.E., Kondejewski,N., Leal,B.,
Lichtarge,O., Liu,W., Logan,O., Lu,J., Martinez,C., Moore,S.,
Moorish,T., Nguyen,N., Oswal,G., Pampall,L.R., Parish,B.J.,
Perez,L.M., Rashid,N.D., Rives,C.M., Scherer,S.E., Shen,H.,
Simon,M.L., Vo,Q.K., Wei,Y., Williamson,A.L., Worley,K., Zhou,X.,
Naylor,S.L., and Gibbs,R.A.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 88326)

REFERENCE
AUTHORS Worley,K.C.
JOURNAL Direct Submission
TITLE Submitted (04-NOV-1998) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 88326)

REFERENCE
AUTHORS Worley,K.C.
JOURNAL Direct Submission
TITLE Submitted (27-FEB-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 88326)

REFERENCE
AUTHORS Worley,K.C.
JOURNAL Direct Submission
TITLE Submitted (12-MAR-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
5 (bases 1 to 88326)

COMMENT Submitted (12-JAN-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Mar 1, 1999 this sequence version replaced gi:4033655.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the
entire insert of this clone. Overlapping regions of clones are only
sequenced and submitted once, so the sequence for the remainder of
the insert may be found in the record for the adjacent clones.
Overlapping clones are noted at the beginning and end of the
Features Listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches
of a local database that includes entries from dbSTS, GDB, and
local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green,
unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
EST and cDNA sequences. Genes demonstrate at least two exons
flanked by consensus splice sites that maintained sequence
continuity across the splice junctions. Sequences that are not
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum
standard of double strand coverage with a minimum of 2 clones and 2
reads with no ambiguities or 2 chemistries with a minimum of 2
clones and 3 reads with no ambiguities. If the sequence quality for
a region does not meet this standard, it will be indicated in the
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
standards - estimated error rate less than 1 per 10,000 bases.
Reports of lowest quality individual bases and measures of base
quality are listed below. Description of the metrics can be found
at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

FEATURES

source

QUALSTAT-REPORT.
Location/Qualifiers

1..88326
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3p21.3-4"
/clone="RP015-751E10"
complement(1..61)
/rpt_family="Alu"
complement(63..364)
/rpt_family="L1ME"
complement(380..671)
/rpt_family="AluX"
complement(684..827)
/rpt_family="L1ME"
complement(877..966)
/rpt_family="L1ME"
1009..1297
/rpt_family="Aluub"
1299..1327
/rpt_family="AT-rich"
1336..1366
/rpt_family="7SLRNA"
1431..1728
/rpt_family="AluSc"
complement(1772..1914)
/rpt_family="L1ME"
complement(2104..2310)
/rpt_family="AluDo"
complement(2356..2650)
/rpt_family="AluY"
complement(2651..2790)
/rpt_family="L1M4"
complement(2793..2958)
/rpt_family="AluX"
complement(2959..3259)
/rpt_family="AluY"
complement(3262..3402)
/rpt_family="AluX"
complement(3407..3533)
/rpt_family="L1M4"
complement(3554..3785)
/rpt_family="AluSg"
complement(3806..3891)
/rpt_family="L1P"
complement(4023..4181)
/rpt_family="L2"
4182..4486
/rpt_family="AluY"
complement(4567..4610)
/rpt_family="L2"
4833..5049
/rpt_family="AluSg/x"
5733..6014
/rpt_family="AluDo"
6341..6659
misc-feature
/note="Region: Clone qk32c09.x1 Homo sapiens cDNA
A1245778"

```
misc_feature      10007..14335
                  /note="Region: Similar to Human mRNA for KIAA0279 gene
                  D87469"
repeat_region     complement(12482..12505)
                  /rpt_family="(CA)n"
gene              join(18532..18996,18755..18896,18994..19114,19398..19582,
                  19892..19994,21255..21352,21545..21754,22012..22208,
                  22477..22632,23545..23711,24216..24383,24605..24731,
                  24824..25030,25264..25438,25645..25757,26178..26257,
                  26796..26919,27390..27506,27589..27737,28006..28115,
                  28464..28623,28949..29149,29914..30799,32243..34004)
                  /gene="Homo sapiens mRNA for MEGF2 AB011536"
repeat_region     23873..23986
                  /rpt_family="(CAGA)n"
repeat_region     complement(26062..26140)
                  /rpt_family="(GA)n"
repeat_region     complement(31427..31535)
                  /rpt_family="L2"
repeat_region     31805..32008
                  /rpt_family="MIR"
misc_feature      34641..34866
                  /note="Region: Homo sapiens CpG island DNA genomic MseI
                  fragment Z58606"
repeat_region     42837..43001
                  /rpt_family="FAM"
gene              join(43002..43597,43788..43842,44119..44255,44473..44750)
                  /gene="Unigene cluster containing AA057543 and AA411587"
                  44592..44724
                  /gene="Unigene cluster containing AA057543 and AA411587"
STS              45136..45393
                  /db_xref="dbSTS:23502"
                  /standard_name="D3S4222"
repeat_region     45136..45393
                  /rpt_family="AluSx"
repeat_region     45339..45633
                  /rpt_family="AluSx"
repeat_region     45674..45879
                  /rpt_family="MIR"
repeat_region     46117..46224
                  /rpt_family="AluSg/X"
repeat_region     complement(46711..46880)
                  /rpt_family="AluJb"
repeat_region     complement(46901..47180)
                  /rpt_family="AluSx"
repeat_region     complement(47189..47498)
                  /rpt_family="AluY"
repeat_region     complement(47499..47782)
                  /rpt_family="AluSg"
misc_feature      47716..47800
                  /note="Sequence is generated from PCR product only"
                  /function="Low coverage"
misc_feature      47802..48101
                  /function="Gap of approximately 300 bps"
repeat_region     complement(48158..48460)
                  /rpt_family="AluJb"
gene              join(48873..49019,49455..49645)
                  /gene="Unigene cluster containing AA292850 and AA405356"
                  49867..50178
                  /rpt_family="AluJb"
repeat_region     complement(50948..51140)

Query Match      0.6%; Score 20; DB 9; Length 88326;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1961 ctacattccagcaggagctg 1980
      |||||||
Db 72992 CTACATTCACGACGCGCTG 73011

RESULT 65
AC084447 88839 bp DNA INV 04-NOV-2000
LOCUS AC084447/c
DEFINITION Caenorhabditis briggsae cosmid CB019G12, complete sequence.
```

```
ACCESSION AC084447
VERSION AC084447.1 GI:11094897
KEYWORDS HMG.
SOURCE Caenorhabditis briggsae.
ORGANISM Caenorhabditis briggsae
REFERENCE 1
AUTHORS Rhabdotox; Metazoa; Nematoda; Chromadorea; Rhabditida;
TITLE Rhabdotox; Rhabditidae; Peloderinae; Caenorhabditis.
JOURNAL Washington University Genome Sequencing Center.
REFERENCE 2
AUTHORS The C. briggsae Genome Sequencing Project
TITLE Unpublished
JOURNAL 2 (bases 1 to 88839)
REFERENCE Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (04-NOV-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA
e-mail: jspleth@wustl.edu

FEATURES
source
Location/Qualifiers
1..88839
/organism="Caenorhabditis briggsae"
/strain="Gujarat G16"
/db_xref="taxon:6238"
/clone="CB019G12"
9330..9401
/note="codon recognized: CCA"
8814..9885
/product="tRNA-Pro"
/note="codon recognized: CCU"
30927..30998
/product="tRNA-Pro"
/note="codon recognized: GGA"
30927..30998
/product="tRNA-Gly"
BASE COUNT 28080 a 16056 c 16645 g 28058 t
ORIGIN

Query Match      0.6%; Score 20; DB 3; Length 88839;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2749 aggtcccaactcgtcaag 2768
      |||||||
Db 67631 AGGTCCCACTCGTCAAG 67612

RESULT 66
AL391278 91733 bp DNA HMG 13-JUN-2001
LOCUS AL391278/c
DEFINITION Homo sapiens chromosome 1 clone RP5-885P2, *** SEQUENCING IN
PROGRESS ***; 29 unordered pieces.
ACCESSION AL391278
VERSION AL391278.10 GI:12331116
KEYWORDS HMG; HTGS-PHASE1; HTGS-CANCELLED.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Mclay, K.
TITLE Direct Submission
JOURNAL Submitted (13-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
```


misc_feature /note="assembly_fragment:00879"
75588.80699 /note="assembly_fragment:00912"
misc_feature 80800.84813 /note="assembly_fragment:00938"
misc_feature 84914.87462 /note="assembly_fragment:00953"
misc_feature 87563.91733 /note="assembly_fragment:00807
clone_end:SP6
vector_side:right"
BASE COUNT 25532 a 18159 c 18530 g 26693 t 2819 others
ORIGIN

Query Match 0.6%; Score 20; DB 2; Length 91733;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2596 attatcagaataaccgt 2615
Db 88326 ATTATCAGAAATACCTG 88307
|||||

RESULT 67
AC073294
LOCUS AC073294 93409 bp DNA HTG 26-MAR-2001
DEFINITION Mus musculus clone CT7-32119 strain 129 SV, WORKING DRAFT SEQUENCE,
1 ordered pieces.
ACCESSION AC073294
VERSION AC073294.1 GI:8493569
KEYWORDS HTG: HTGS_PHASE2; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 93409)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Mouse
JOURNAL Unpublished
2 (bases 1 to 93409)
REFERENCE DOE Joint Genome Institute.
AUTHORS Direct Submission
TITLE Submitted (13-JUN-2000) Production Sequencing Facility, DOE Joint
JOURNAL Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT -----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

FEATURES
source
1. 93409
/organism="Mus musculus"
/strain="129 SV"
/db_xref="taxon:10090"
/clone="CT7-32119"
/clone_lib="CltDCJ7 mouse BAC library"
BASE COUNT 25162 a 21194 c 20677 g 25391 t 985 others
ORIGIN

Query Match 0.6%; Score 20; DB 2; Length 93409;
Best Local Similarity 100.0%; Pred. No. 76;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2172 aacattatctctcatggt 2191
Db 76311 AACATTATCTCTCATGCT 76330
|||||

RESULT 68
AC083819/c
LOCUS AC083819/c 102165 bp DNA HTG 06-NOV-2000
DEFINITION Mus musculus chromosome 1 clone RP23-285F20, *** SEQUENCING IN
PROGRESS ***; 7 unordered pieces.
ACCESSION AC083819
VERSION AC083819.4 GI:10803522
KEYWORDS HTG: HTGS_PHASE1.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 102165)
AUTHORS Metzker, M.L., Lewis, L.R., Hume, J., Edwards, C., Harris, C.,
Dederich, D., Thomas, S., Okwunonu, G., Carllock, C., Garner, T.,
Addison, S., Pace, A., Williams, G., Bonnin, D., Brooks, A., Brown, J.,
Buhay, C., Bunac, C., Burkett, C., Chacko, J., Chen, G., Chen, Z.,
Cox, C., Davis, C., Delgado, O., Ding, Y., Dugan-Rocha, S.,
Fernandez, C., Ferraguto, D., Forcum-Fansey, J., Gill, R.,
Gorrell, J.H., Gunaratne, P., Haller, G., Hernandez, J., Hognes, M.,
Hosak, H., Hou, X., Huber, J., Jackson, L., Jia, Y., Kelly, J., Kelly, S.,
Kovar, C., Liu, J., Liu, W., Louisgied, H., Lozado, R.J., Martin, R.,
Massey, E., McLeod, M.P., Mei, G., Moore, S., Morgan, M., Morris, S.,
Neal, D., Nelson, A., Nguyen, R., Nguyen, N., Ogih, M., Parish, B.,
Perez, L., Reiter, D., Say, J., Shen, H., Vasquez, L., Wellington, S.,
Williamson, A., Wrenstord, G., Zhou, X., Bouck, J., Hodgson, A.,
Muzny, D.M., Rives, M., Scherer, S., Sodergren, E., Weinstein, G.,
Worley, K., and Gibbs, R.
TITLE Direct Submission
JOURNAL Unpublished
2 (bases 1 to 102165)
REFERENCE Worley, K.C.
AUTHORS Direct Submission
TITLE Submitted (01-OCT-2000) Human Genome Sequencing Center, Department
JOURNAL of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
COMMENT On Oct 16, 2000 this sequence version replaced gi:10801956.
-----Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
-----Project Information
Center project name: MATO
Center clone name: RP23-285F20
-----Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-primer Bodipy; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 95179 bases at least Q40
Consensus quality: 98007 bases at least Q30
Consensus quality: 99497 bases at least Q20
Estimated insert size: 100390; sum-of-contrs estimation
Quality coverage: 0x in Q20 bases; agrose-gp estimation
Quality coverage: 2.9x in Q20 bases; sum-of-contrs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will

	Location	Qualifiers
*	1	24502: contig of 24502 bp in length
*	24503	24502: gap of unknown length
*	24603	40809: contig of 16207 bp in length
*	40810	40809: gap of unknown length
*	40910	63388: contig of 22480 bp in length
*	63390	63489: gap of unknown length
*	63390	78501: contig of 15012 bp in length
*	78502	78601: gap of unknown length
*	78602	89887: contig of 1186 bp in length
*	89788	89887: gap of unknown length
*	89888	96681: contig of 6794 bp in length
*	96682	96781: gap of unknown length
*	96782	102165: contig of 5384 bp in length

FEATURES	source	Location/Qualifiers
		1. 102165
		/organism="Mus musculus"
		/db_xref="taxon:10090"
		/chromosome="1"
		/clone="RP3-885E20"
BASE COUNT	30175	a 21419 c 20896 g 29070 t 605 others
ORIGIN		

Query Match	0.6%	Score	20	DB	2	Length	102165
Best Local Similarity	100.0%	Pred. No.	76				
Matches	20	Conservative	0	Mismatches	0	Indels	0
						Gaps	0

QY	2055	aagcaagatatcacatatct	2074
Db	51969	AAGCAAGATATCACATATCT	51950

RESULT	69		
AL557352			
LOCUS	AL557352	106601 bp	DNA
DEFINITION	Human DNA sequence from clone Rpl1-799G19 on chromosome 6, complete sequence.		PRI 20-NOV-2000
ACCESSION	AL557352.11	GI:11321993	
VERSION			
KEYWORDS	HTG.		
SOURCE	human.		

REFERENCE 1 (bases 1 to 106601)
AUTHORS Phillipmore, B.
TITLE Direct Submission
JOURNAL Submitted (19-NOV-2000) Sanger Centre, Hinxton, Cambridgeshire,

COMMENT

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at

chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/chr6>
RP11-799G19 is from the library RPc1-11.3 constructed at the

Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/vectors/pBac3.6>
IMPORTANT: This sequence is not the entire insert of clone RP11-799G19. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone RP11-799G19 is at 1 in this sequence. The true left end of clone RP11-532N4 is at 106502 in this sequence.

location/Qualifiers
 1..106601 ..

```
FEATURES                                Location/Qualifiers
source                                  1..106601
                                         /organism="Homo sapiens"
                                         /db_xref="taxon:9606"
                                         /chromosome="6"
                                         /clone="RP11-796G19"
                                         /clone_lib="RPCI-11.3"
                                         8..506
misc_feature                           /note="match: GSS: Em:AQ521292"
                                        961..2386
repeat_region                          /note="TIGER2 repeat: matches 1..1483 of consensus"
                                        2387..2755
                                        /note="THEIC repeat: matches 1..371 of consensus"
                                        2756..4009
                                        /note="TIGER2 repeat: matches 1483..2718 of consensus"
repeat_region                          4387..4995
                                        /note="LIP3 repeat: matches 768..1393 of consensus"
                                        4995..6352
                                        /note="LIP12 repeat: matches 4779..6163 of consensus"
                                        9015..9156
                                        /note="L2 repeat: matches 2608..2750 of consensus"
                                        9690..9802
                                        /note="MIR repeat: matches 56..172 of consensus"
                                        10378..10458
                                        /note="MIR repeat: matches 71..153 of consensus"
                                        10787..10832
                                        /note="23 copies 2 mer at 78% conserved"
                                        11644..11807
                                        /note="L2 repeat: matches 2581..2749 of consensus"
                                        11817..11984
                                        /note="MIR repeat: matches 5..186 of consensus"
                                        12435..12545
                                        /note="MIR repeat: matches 8..128 of consensus"
                                        13942..13995
                                        /note="18 copies 3 mer tta 88% conserved"
                                        13954..13998
                                        /note="3 copies 15 mer 93% conserved"
                                        13896..14278
                                        /note="AluY repeat: matches 1..286 of consensus"
                                        14302..14364
                                        /note="L2 repeat: matches 2634..2701 of consensus"
                                        14716..14753
                                        /note="19 copies 2 mer ac 81% conserved"
                                        15675..15806
                                        /note="FLAM_A repeat: matches 1..133 of consensus"
                                        17568..17732
                                        /note="L2 repeat: matches 1683..1850 of consensus"
                                        17850..18255
                                        /note="match: GSS: Em:AQ790581"
                                        19598..19685
                                        /note="MLTJ1 repeat: matches 221..337 of consensus"
                                        20418..20555
                                        /note="MER5A repeat: matches 47..182 of consensus"
                                        29307..29439
                                        /note="MIR repeat: matches 87..245 of consensus"
                                        29675..30120
                                        /note="MTRF repeat: matches 91..529 of consensus"
                                        30350..30644
                                        /note="Alusq repeat: matches 1..313 of consensus"
                                        30687..30860
                                        /note="L2 repeat: matches 1728..1906 of consensus"
                                        33081..33579
                                        /note="match: GSS: Em:AQ783759"
                                        33928..34076
```

```

/note="MIR repeat: matches 4. .165 of consensus"
34988. .35085
/note="LIP3 repeat: matches 6048. .6146 of consensus"
35092. .35173
/note="LIMC2 repeat: matches 6245. .6328 of consensus"
35174. .35517
/note="MLT1A1 repeat: matches 1. .354 of consensus"
35525. .36385
/note="LIMC1 repeat: matches 5357. .6250 of consensus"
36388. .37713
/note="LIM4 repeat: matches 3807. .5147 of consensus"
40703. .40763
/note="L2 repeat: matches 2676. .2734 of consensus"
40918. .41362
/note="match: GSS: Em:AQ756546"
42175. .42385
/note="match: GSS: Em:AQ113333"
/note="match(42487. .42933)
/note="match: GSS: Em:AQ188019"
complement(42506. .42945)
/note="match: GSS: Em:AQ408008"
complement(43648. .44149)
/note="match: GSS: Em:AQ563645"
complement(43831. .44224)
/note="match: GSS: Em:A2335395"
44050. .44510
/note="match: GSS: Em:AQ117470"
44201. .44518
/note="match: GSS: Em:AQ710247"
45984. .46189
/note="LIM4C repeat: matches 1430. .1645 of consensus"
46033. .46435
/note="match: GSS: Em:AQ622979"
46036. .46436
/note="match: GSS: Em:AQ622990"
46385. .46820
/note="LIM1 repeat: matches 4942. .5389 of consensus"
46829. .47635
/note="LIMC1 repeat: matches 3969. .4776 of consensus"
47842. .47889
/note="24 copies 2 mer ta 97% conserved"
47844. .47897
/note="3 copies 18 mer 85% conserved"
47890. .48044
/note="LIP repeat: matches 3526. .3683 of consensus"
48082. .48365
/note="LIM4 repeat: matches 3095. .3390 of consensus"
49813. .50122
/note="AluSg repeat: matches 1. .295 of consensus"
50653. .50930
/note="L2 repeat: matches 2008. .2286 of consensus"
51027. .51334
/note="AluY repeat: matches 1. .308 of consensus"
51368. .51496
/note="LIP1A10 repeat: matches 6029. .6156 of consensus"
51510. .51947
/note="L2 repeat: matches 1292. .1752 of consensus"
52009. .52098
/note="45 copies 2 mer ga 84% conserved"
52028. .52099
/note="4 copies 18 mer 91% conserved"
52282. .52518
/note="LIP27 repeat: matches 5559. .5793 of consensus"
52519. .52860
/note="LIP47 repeat: matches 5789. .6143 of consensus"
53493. .54159
/note="match: GSS: Em:B93480"
53511. .53921
/note="match: GSS: Em:AQ011862"
55163. .55556
/note="MER57B repeat: matches 1. .403 of consensus"
55697. .56012
/note="AluSc repeat: matches 1. .305 of consensus"

```

```

misc_feature 56584. .57061
/note="match: GSS: Em:AQ887852"
56588. .56963
misc_feature 56588. .56963
/note="match: GSS: Em:AQ84739"
56874. .56931
repeat_region 56874. .56931
/note="29 copies 2 mer gt 70% conserved"
57865. .58157
repeat_region 57865. .58157
/note="AluSg repeat: matches 1. .293 of consensus"
58991. .59162
repeat_region 58991. .59162
/note="MER5A repeat: matches 6. .189 of consensus"
59736. .59927
repeat_region 59736. .59927
/note="MER5B repeat: matches 1. .177 of consensus"
59928. .60263
repeat_region 59928. .60263
/note="L2 repeat: matches 144. .499 of consensus"
60293. .60745

```

```

Query Match 0.6%; Score 20; DB 9; Length 106601;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 306 catcacatcacagaagaga 325
Db 58795 CATCAGACATCAGAGAGA 58814

```

```

RESULT 70
AL139036/C AL139036 109810 bp DNA PRI 26-MAY-2001
LOCUS Human DNA sequence from clone RP11-285H24 on chromosome
DEFINITION 13q22.1-22.3, complete sequence.
ACCESSION AL139036
VERSION AL139036.12 GI:14270405
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens

```

```

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 109810)
TITLE Direct Submission
JOURNAL Submitted (26-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humbugys@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On May 31, 2001 this sequence version replaced gi:14269884.

```

COMMENT

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at <http://www.sanger.ac.uk/Projects/Celegans/wormpep>

This sequence was generated from part of bacterial clone contigs of human chromosome 13, constructed by the Sanger Centre Chromosome 13 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr13>

RP11-285H24 is from the library RPI1-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-285H24. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true left end of clone RP11-810G21 is at 109711 in this sequence. The true right end of clone RP11-157H4 is at 100 in this sequence.

```
FEATURES
source          Location/Qualifiers
                1..109810
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="13"
                /map="q22.1-22.3"
                /clone="RP11-285H24"
                /clone.lib="RPC1-11.1"
                complement(1..98)
                /note="match: GSS: Em:AQ374965"
                205..514
                /note="match: GSS: Em:AQ092767"
                1091..1201
                /note="MLT1H repeat: matches 81..189 of consensus"
                1395..1522
                /note="MLT1H repeat: matches 409..510 of consensus"
                2642..3032
                /note="match: GSS: Em:AQ477479"
                2718..3197
                /note="MLT1D repeat: matches 1..502 of consensus"
                3752..3940
                /note="7 copies 27 mer 66% conserved"
                3773..4126
                /note="59 copies 6 mer 9999ga 54% conserved"
                3817..4088
                /note="16 copies 17 mer 59% conserved"
                3818..4076
                /note="7 copies 37 mer 63% conserved"
                5037..5248
                /note="MIR repeat: matches 12..248 of consensus"
                complement(6151..6886)
                /note="match: GSS: Em:AZ519175"
                complement(6237..6883)
                /note="match: GSS: Em:AQ283095"
                complement(6380..6889)
                /note="match: GSS: Em:AQ581403"
                6623..6699
                /note="HAL1 repeat: matches 669..746 of consensus"
                7701..7870
                /note="MIR repeat: matches 60..220 of consensus"
                8648..8683
                /note="18 copies 2 mer tt 80% conserved"
                complement(9052..9485)
                /note="match: GSS: Em:AQ620944"
                9149..9432
                /note="Alusx repeat: matches 1..284 of consensus"
                9905..10103
                /note="MER63A repeat: matches 1..210 of consensus"
                10116..10217
                /note="L1ME3A repeat: matches 6016..6116 of consensus"
                11190..11776
                /note="match: GSS: Em:AZ374306"
                11196..11815
                /note="match: GSS: Em:AZ374726"
                13390..13554
                /note="MIR repeat: matches 12..168 of consensus"
                15651..15963
                /note="AluY repeat: matches 1..306 of consensus"
                18346..18991
                /note="match: GSS: Em:AZ516011"
                18347..18878
                /note="match: GSS: Em:AQ384942"
                18568..18864
                /note="AluDb repeat: matches 3..306 of consensus"
                complement(18803..19250)
                /note="match: GSS: Em:AQ801630"
                20484..20699
                /note="MIR repeat: matches 11..252 of consensus"
                23381..23476
                /note="L1M4 repeat: matches 4013..4108 of consensus"
```

```
repeat_region 23477..23773
                /note="AluY repeat: matches 1..299 of consensus"
repeat_region 23774..24043
                /note="L1M4 repeat: matches 4107..4379 of consensus"
                24092..24357
                /note="L1ME3 repeat: matches 5861..6131 of consensus"
                complement(24264..24689)
                /note="match: GSS: Em:AQ451631"
                24793..24874
                /note="MIR repeat: matches 67..145 of consensus"
                25008..25518
                /note="match: GSS: Em:AQ517147"
                25611..25698
                /note="MIR repeat: matches 2..91 of consensus"
                26142..26439
                /note="Alusx repeat: matches 3..307 of consensus"
                27548..27704
                /note="MER33 repeat: matches 22..166 of consensus"
                27705..28031
                /note="MER2 repeat: matches 4..345 of consensus"
                28032..28156
                /note="MER33 repeat: matches 166..324 of consensus"
                30049..30177
                /note="FLM_A repeat: matches 1..133 of consensus"
                31105..31202
                /note="MIR repeat: matches 12..111 of consensus"
                31520..31675
                /note="MIR repeat: matches 100..261 of consensus"
                32359..32560
                /note="L1M4 repeat: matches 5318..5518 of consensus"
                32561..32861
                /note="AluDb repeat: matches 1..296 of consensus"
                32862..33218
                /note="L1M4 repeat: matches 5518..5905 of consensus"
                33290..33388
                /note="L1ME3A repeat: matches 6029..6129 of consensus"
                33458..33489
                /note="16 copies 2 mer aa 100% conserved"
                33902..33951
                /note="25 copies 2 mer at 74% conserved"
                34308..34780
                /note="L1M48 repeat: matches 5785..6288 of consensus"
                35285..35684
                /note="match: GSS: Em:AQ064667"
                36808..37104
                /note="AluY repeat: matches 1..296 of consensus"
                37255..37308
                /note="2 copies 27 mer 94% conserved"
                37728..37913
                /note="MER63A repeat: matches 1..210 of consensus"
                38654..38727
                /note="2 copies 37 mer 100% conserved"
                39086..39226
                /note="MER3B repeat: matches 29..158 of consensus"
                39258..39424
                /note="MER3B repeat: matches 2..172 of consensus"
                39564..39981
                /note="MLT1C repeat: matches 17..466 of consensus"
                44189..44535
                /note="L1MC/D repeat: matches 5436..5813 of consensus"
                45138..45321
                /note="MIR repeat: matches 81..260 of consensus"
                49524..49727
                /note="MIR repeat: matches 24..252 of consensus"
                49821..50326
                /note="match: GSS: Em:AQ169074"
                49836..50027
                /note="L1MC/D repeat: matches 5611..5813 of consensus"
                50250..50567
                /note="Alusx repeat: matches 1..312 of consensus"
                51288..51497
                /note="MER2 repeat: matches 1..209 of consensus"
                52175..52231
```

repeat_region /note="MIR repeat: matches 15..74 of consensus"
53123..53421
/note="AlusJ1 repeat: matches 1..299 of consensus"
repeat_region 54142..54423
/note="L1MD2 repeat: matches 6011..6278 of consensus"
misc_feature 54454..55060
/note="match: GSS: Em:AQ195363"
repeat_region 54535..54652

Query Match 0.6%; Score 20; DB 9; Length 109810;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2826 ttctggaagaccctcgaa 2845
|||||
Db 62720 TTTCGAAAGAACCTCTGAA 62701

RESULT 71
LOCUS AC010902 114169 bp DNA PRI 25-MAR-2001
DEFINITION Homo sapiens clone Rpl1-549J7, complete sequence.
ACCESSION AC010902
VERSION AC010902.4 GI:10440867
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 114169)
AUTHORS Waterston,R.H.
JOURNAL Unpublished
TITLE The sequence of Homo sapiens clone
AUTHORS 2 (bases 1 to 114169)
JOURNAL Waterston,R.H.
REFERENCE Direct Submission
TITLE Submitted (25-SEP-1999) Genome Sequencing Center, Washington
JOURNAL University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
3 (bases 1 to 114169)
AUTHORS Waterston,R.
JOURNAL Direct Submission
TITLE Submitted (30-SEP-2000) Department of Genetics, Washington
JOURNAL University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
4 (bases 1 to 114169)
AUTHORS Waterston,R.H.
JOURNAL Direct Submission
TITLE Submitted (25-MAR-2001) Genome Sequencing Center, Washington
JOURNAL University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Sep 30, 2000 this sequence version replaced gi:7631065.
Center Project name: H_NH0549J07.
Location/Qualifiers
FEATURES
source 1..114169
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Rpl1-549J7"
BASE COUNT 34392 a 22524 c 23021 g 34232 t
ORIGIN

Query Match 0.6%; Score 20; DB 9; Length 114169;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 295 aaagcttcttcacagaca 314
|||||
Db 74191 AAAGCTTTTCATCAGACA 74172

RESULT 72
LOCUS AC022124/c 120733 bp DNA PRI 21-JUL-2001

DEFINITION Homo sapiens chromosome 5 clone CTD-2075G19, complete sequence.
ACCESSION AC022124
VERSION AC022124.6 GI:14993687
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 120733)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL Direct Submission
TITLE Unpublished
REFERENCE 2 (bases 1 to 120733)
AUTHORS DOE Joint Genome Institute.
JOURNAL Direct Submission
TITLE Submitted (26-JAN-2000) Production Sequencing Facility, DOE Joint
JOURNAL Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 120733)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL Direct Submission
TITLE Submitted (01-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell
JOURNAL Drive, Walnut Creek, CA 94598, USA
4 (bases 1 to 120733)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL Direct Submission
TITLE Submitted (21-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell
JOURNAL Drive, Walnut Creek, CA 94598, USA
On Jul 21, 2001 this sequence version replaced gi:14277276.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.7% of Sequence;
Estimated Total Number of Errors is 0.3.
Location/Qualifiers
FEATURES
source 1..120733
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-2075G19"
BASE COUNT 39434 a 20042 c 22104 g 39153 t
ORIGIN

Query Match 0.6%; Score 20; DB 9; Length 120733;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 400 aagatattgacattatltt 419
|||||
Db 61200 AAGATATTGACATTATTTT 61181

RESULT 73
LOCUS AC008390 121931 bp DNA PRI 20-DEC-2000
DEFINITION Homo sapiens chromosome 5 clone CTC-235M3, complete sequence.
ACCESSION AC008390
VERSION AC008390.7 GI:11908269
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 121931)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL Direct Submission
TITLE Unpublished
REFERENCE 2 (bases 1 to 121931)
AUTHORS DOE Joint Genome Institute.
JOURNAL Direct Submission
TITLE Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
JOURNAL Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

```

REFERENCE      3 (bases 1 to 121931)
AUTHORS        DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE          Direct Submission
JOURNAL        Submitted (20-DEC-2000) DOE Joint Genome Institute, 2800 Mitchell
                Drive, Walnut Creek, CA 94598, USA
                On Dec 20, 2000 this sequence version replaced gi:7711258.
COMMENT        Draft Sequence produced by DOE Joint Genome Institute
                www.jgi.doe.gov
                Finishing Completed at Stanford Human Genome Center
                www.shgc.stanford.edu
                Quality: Phrap Quality >=40 99.8% of Sequence;
                Estimated Total Number of Errors is 0.2.

FEATURES
  source
    1..121931
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="5"
      /clone="CTC-235M3"

BASE COUNT     39800 a 20269 c 22375 g 39487 t

ORIGIN
Query Match    0.6%; Score 20; DB 9; Length 121931.
Best Local Similarity 100.0%; Pred. No. 75;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 400 aagatattgacattatttt 419
    |||||||
Db 62415 AAGATATTGACATATTATTTT 62396

RESULT 74
AC010072/2c    124347 bp    DNA    PRI    19-NOV-1999
DEFINITION     Homo sapiens chromosome 14q31 clone CTD-217314 containing TSHR
ACCESSION      AC010072
VERSION        AC010072.5 GI:6453843
KEYWORDS       HTG.
SOURCE         human.
ORGANISM       Homo sapiens
                Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE      1 (bases 1 to 124347)
AUTHORS        Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B.,
                Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G.,
                James,R., Kaur,A., Madan,A., Owen,M.P., Ratcliffe,A., Shaffer,T.
                and Hood,L.
                Sequencing of human chromosome 14
                Unpublished
                2 (bases 1 to 124347)
                Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B.,
                Bloom,S., Dors,M., Dickhoff,R., Harrison,G., James,R., Lasky,S.,
                Madan,A., Ratcliffe,A., Shaffer,T. and Hood,L.
                Direct Submission
                Submitted (11-SEP-1999) Multimegabase Sequencing Center, University
                of Washington, PO BOX 357730, Seattle, WA 98195, USA
                3 (bases 1 to 124347)
                Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B.,
                Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G.,
                James,R., Kaur,A., Madan,A., Owen,M.P., Ratcliffe,A., Shaffer,T.
                and Hood,L.
                Direct Submission
                Submitted (19-NOV-1999) Multimegabase Sequencing Center, University
                of Washington, PO BOX 357730, Seattle, WA 98195, USA
                On Nov 19, 1999 this sequence version replaced gi:6114900.
COMMENT        ----- Genome Center
                Center: Multimegabase Sequencing Center
                Center code: UWMSC
                Web site: http://chrtona.mbt.washington.edu/msg_www
                Contact: leerowen@u.washington.edu
                ----- Summary Statistics
                Sequencing vector: pUC18; L08752

FEATURES
  source
    1..124347
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="14"
      /map="14q31"
      /clone="CTD-217314"
      /clone_id="Caltech human BAC library D"
      /note="this clone overlaps RP11-114N19, Accession
      AC007262"
      complement(join(42575..42673,43714..43786,47934..48025,
      52809..52927,53951..54077,62318..62371))
      /note="Intron-exon boundaries defined in relation to EST
      A154064, the closest BLASTX similarity is to a
      drosophila potassium channel protein"
      /codon_start=1
      /product="unknown"
      /protein_id="AAF09033.1"
      /db_xref="GI:6453845"
      /translation="MIGRYREYNGCGAATPEIKESLFOSTIDLRSGRLINSGRST
      SVTSLASADLDGGTSELHHPPTPLKDYCDPOGIKMRKRTGVRFVQETDPTOLH
      GFHSLRDLSSSEDIRLGDFNRELSRKRSDAEKRALELTETKLENAQKQEVFVNS
      LGKVKIQCLSSAEVDNVPPIVWOCGI"
      95810..95860
      /note="low quality data"
      103634..>103846
      /note="this is the 5' end of TSHR, found in M73745"
      /product="thyroid stimulating hormone receptor"
      103677..>103846
      /note="thyroid stimulating hormone receptor"
      /codon_start=1
      /product="TSHR"
      /protein_id="AAF09032.1"
      /db_xref="GI:6453844"
      /translation="MRPADLQVLVLDLPRLDGMGCGSSPCECHQEDPFRVTKXDI
      ORIPSLPSTQTL"
      109619..124347
      /note="Overlap with RP11-114N19, Accession AC007262"
      110216
      /note="217314: c; 114N19: t"
      /replace="t"
      110755
      /note="217314: g; 114N19: a"
      /replace="a"
      111115
      /note="217314: c; 114N19: g"
      /replace="g"
      114198
      /note="217314: g; 114N19: a"
      /replace="a"
      114690
      /note="217314: c; 114N19: t"
      /replace="t"
      115183
      /note="217314: t; 114N19: tatat"
      /replace="tatat"
      115210..115280
      /note="low quality data"
      116847
      /note="217314: t; 114N19: c"
      /replace="c"
      117131
      /note="217314: g; 114N19: a"
      /replace="a"
      117408
      /note="217314: a; 114N19: c"
      /replace="c"
      117633
      /note="217314: g; 114N19: t"
      /replace="t"

```

Chemistry: Big Dye terminators and primers
 Assembly program: Phrap; version 0.990399

```

variation      118901
                /note="217314: a; 114N19: g"
                /replace="g"
variation      118988
                /note="217314: c; 114N19: t"
                /replace="t"
variation      119250
                /note="217314: a; 114N19: g"
                /replace="g"
variation      119718
                /note="217314: g; 114N19: t"
                /replace="t"
variation      120914
                /note="217314: c; 114N19: a"
                /replace="a"
variation      121395
                /note="217314: caaaaaaaaaaaaaaa; 114N19: c"
                /replace="c"
variation      123019
                /note="217314: g; 114N19: a"
                /replace="a"
variation      123721
                /note="217314: c; 114N19: t"
                /replace="t"
variation      123900
                /note="217314: t; 114N19: c"
                /replace="c"
variation      123926
                /note="217314: t; 114N19: t"
                /replace="t"
variation      124015
                /note="217314: t; 114N19: c"
                /replace="c"
variation      124182
                /note="217314: g; 114N19: a"
                /replace="a"
BASE COUNT    43520 a 24500 c 22595 g 33732 t
ORIGIN
Query Match   0.6%; Score 20; DB 9; Length 124347;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 405 attgacattatcttact 424
|||||
Db 99037 ATTGACATTATTTTACTT 99018

```

```

RESULT 75
AC004829
LOCUS          AC004829 129837 bp DNA PRI 21-DEC-1999
DEFINITION    Homo sapiens PAC clone RP4-530J23 from 7q21.1-q21.3, complete
                sequence.
ACCESSION     AC004829
VERSION       AC004829.2 GI:4309818
KEYWORDS      HTG.
SOURCE        human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE     1 (bases 1 to 129837)
AUTHORS      Sulston, J.E. and Waterston, R.
TITLE        Toward a complete human genome sequence
JOURNAL       Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE       99063792
REFERENCE     2 (bases 1 to 129837)
AUTHORS      Harmon, G., Langston, Y. and Kellen, J.
TITLE        The sequence of Homo sapiens PAC clone RP4-530J23
JOURNAL       Unpublished
REFERENCE     3 (bases 1 to 129837)
AUTHORS      Waterston, R.H.
TITLE        Direct Submission

```

```

JOURNAL        Submitted (12-JUN-1998) Genome Sequencing Center, Washington
                University School of Medicine, 4444 Forest Park Parkway, St. Louis,
                MO 63108, USA
REFERENCE      4 (bases 1 to 129837)
AUTHORS       Waterston, R.
TITLE         Direct Submission
JOURNAL       Submitted (25-FEB-1999) Department of Genetics, Washington
                University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE     5 (bases 1 to 129837)
AUTHORS       Waterston, R.
TITLE         Direct Submission
JOURNAL       Submitted (21-DEC-1999) Department of Genetics, Washington
                University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT       On Mar 1, 1999 this sequence version replaced gi:3213168.

----- Genome Center.
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
-----
Center project name: H_DJ0530J23
-----

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and
sequencing collaboration between the NHGRI Chromosome 7 Mapping
Project (Eric D. Green, Director), John D. McPherson in the
Department of Genetics (Washington University), and the Washington
University Genome Sequencing Center. For additional information
about the map position of this sequence, see
http://www.nhgri.nih.gov/DIR/GRB/CHR7, send
mailto:egreen@nhgri.nih.gov, or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
This clone was derived from human PAC library RPCI-4, prepared by
Pietier de Jong and coworkers at the Roswell Park Cancer Institute
(http://bacpac.med.buffalo.edu) using the method described by
Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from
one male donor.
The clone may be obtained either from Genome Systems, Inc.
(http://www.genomesystems.com) or Research Genetics, Inc.
(http://www.resgen.com); or from Pietier de Jong.
VECTOR: pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the right is GSI-67A24, 200 bp overlap.
Actual start of this clone is at base position 1 of RP4-530J23;
actual end is at 7592 of GSI-67A24.

FEATURES
source
1. 129837
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q21.1-q21.3"
/clone="RP4-530J23"
/location.lib="RPCI-4"
1. 1172
/rpt_family="MaLR"
1173..1541
/rpt_family="MaLR"
1542..1962
repeat_region
repeat_region
repeat_region
repeat_region

```

```

repeat_region /rpl_family="L1"
1980. .2394
repeat_region /rpl_family="L1"
2396. .2582
repeat_region /rpl_family="Alu"
2595. .3425
repeat_region /rpl_family="L1"
3426. .3736
repeat_region /rpl_family="Alu"
3757. .3931
repeat_region /rpl_family="L1"
3932. .4095
repeat_region /rpl_family="L1"
4365. .4762
repeat_region /rpl_family="Other"
5079. .5262
repeat_region /rpl_family="Alu"
5332. .5358
repeat_region /rpl_family="MER4-group"
5341. .5804
repeat_region /rpl_family="MER4-group"
6123. .6217
repeat_region /rpl_family="AT_rich"
6810. .6911
repeat_region /rpl_family="AT_rich"
6932. .6960
repeat_region /rpl_family="AT_rich"
6966. .7050
repeat_region /rpl_family="MER73-group"
7086. .7423
repeat_region /rpl_family="L2"
8602. .8649
repeat_region /rpl_family="L2"
8608. .8687
repeat_region /rpl_family="MIR"
8688. .8924
repeat_region /rpl_family="Alu"
9006. .9146
repeat_region /rpl_family="L2"
9147. .9927
repeat_region /rpl_family="MER73-group"
9928. .10346
repeat_region /rpl_family="L2"
11104. .11413
repeat_region /rpl_family="Alu"
complement(12342. .12525)
/db_xref="GI:1113437"
13093. .13412
repeat_region /rpl_family="L1"
14362. .14549
repeat_region /rpl_family="L1"
14553. .14967
repeat_region /rpl_family="L1"
15716. .15747
repeat_region /rpl_family="AT_rich"
15786. .15821
repeat_region /rpl_family="POLY_A"
17133. .17361
repeat_region /rpl_family="MIR"
17455. .17841
repeat_region /rpl_family="MALR"
17962. .18047
repeat_region /rpl_family="L2"
18232. .18295
repeat_region /rpl_family="L1"
18537. .18650
repeat_region /rpl_family="MIR"
18651. .18943
repeat_region /rpl_family="Alu"
18944. .18957
repeat_region /rpl_family="MIR"
18958. .19010
repeat_region /rpl_family="(TA)n"

```

```

repeat_region 19011. .19068
repeat_region /rpl_family="MIR"
20552. .20660
repeat_region /rpl_family="L2"
20891. .21251
repeat_region /rpl_family="MALR"
22154. .22270
repeat_region /rpl_family="L2"
22376. .22531
repeat_region /rpl_family="L2"
22573. .22637
repeat_region /rpl_family="AT_rich"
23327. .23702
repeat_region /rpl_family="MER1_type?"
24527. .24650
repeat_region /rpl_family="MER94"
24950. .25256
repeat_region /rpl_family="Alu"
25634. .25677
repeat_region /rpl_family="(CA)n"
25830. .25999
repeat_region /rpl_family="MIR"
26413. .26449
repeat_region /rpl_family="AT_rich"
26664. .26697
repeat_region /rpl_family="(TA)n"
27087. .27107
repeat_region /rpl_family="AT_rich"
29530. .29577
repeat_region /rpl_family="G-rich"
33727. .33882
repeat_region /rpl_family="MIR"

```

```

Query Match 0.6%; Score 20; DB 9; Length 129837;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 728 attcatggcatgctgctga 747
|||||
Db 128831 ATTGATGGCAGCTGCTGA 128850

```

Search completed: March 25, 2002, 19:03:19
Job time: 26263 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2002, 13:24:21 ; Search time 97.89 Seconds
(without alignments)
7248.499 Million cell updates/sec

File: US-09-697-089-1
Perfect score: 3133
Sequence: 1 cgcctacgcccgcgtggaag.....aagtgactcgaacagcta 3133

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 351203 seqs, 113238999 residues

Word size : 0

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database : Issued Patents_NA:*

- 1: /cgn2_6/ptodata/2/lna/5A.COMB.seq:*
- 2: /cgn2_6/ptodata/2/lna/5B.COMB.seq:*
- 3: /cgn2_6/ptodata/2/lna/6A.COMB.seq:*
- 4: /cgn2_6/ptodata/2/lna/6B.COMB.seq:*
- 5: /cgn2_6/ptodata/2/lna/6C.COMB.seq:*
- 6: /cgn2_6/ptodata/2/lna/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18	0.6	1094	2 US-08-870-518-34	Sequence 34, Appl
2	18	0.6	1515	3 US-08-747-221B-16	Sequence 16, Appl
3	18	0.6	1515	3 US-08-747-221B-17	Sequence 17, Appl
4	18	0.6	1515	4 US-09-005-051-16	Sequence 16, Appl
5	18	0.6	1515	4 US-09-005-051-17	Sequence 17, Appl
6	18	0.6	1611	6 5213972-6	Patent No. 5213972
7	18	0.6	1650	3 US-08-747-221B-60	Sequence 60, Appl
8	18	0.6	1650	3 US-08-747-221B-61	Sequence 61, Appl
9	18	0.6	1650	4 US-09-005-051-60	Sequence 60, Appl
10	18	0.6	1650	4 US-09-005-051-61	Sequence 61, Appl
11	18	0.6	1721	3 US-08-857-213-2	Sequence 2, Appl
12	18	0.6	1600	1 US-08-139-937-11	Sequence 11, Appl
13	18	0.6	1800	5 PCT-US93-11310-11	Sequence 11, Appl
14	18	0.6	1982	3 US-08-747-221B-13	Sequence 13, Appl
15	18	0.6	1982	3 US-08-747-221B-15	Sequence 15, Appl
16	18	0.6	1982	4 US-09-005-051-13	Sequence 13, Appl
17	18	0.6	1982	4 US-09-005-051-15	Sequence 15, Appl
18	18	0.6	2144	3 US-08-747-221B-57	Sequence 57, Appl
19	18	0.6	2144	3 US-08-747-221B-59	Sequence 59, Appl
20	18	0.6	2144	4 US-09-005-051-57	Sequence 57, Appl
21	18	0.6	2144	4 US-09-005-051-59	Sequence 59, Appl
22	18	0.6	3000	4 US-09-192-104-1	Sequence 1, Appl
23	18	0.6	3000	4 US-09-543-446-1	Sequence 1, Appl
24	18	0.6	4982	3 US-08-699-103B-1	Sequence 1, Appl
25	17	0.5	290	4 US-09-037-990B-27	Sequence 27, Appl
26	17	0.5	383	4 US-08-991-789A-171	Sequence 171, Appl
27	17	0.5	661	4 US-09-037-990B-8	Sequence 8, Appl

28	17	0.5	685	4 US-08-998-416-818	Sequence 818, App
29	17	0.5	1536	4 US-09-318-443-5	Sequence 5, Appl
30	17	0.5	1563	2 US-08-743-637B-183	Sequence 183, App
31	17	0.5	1682	4 US-09-318-443-7	Sequence 7, Appl
32	17	0.5	1965	4 US-09-178-252-26	Sequence 26, Appl
33	17	0.5	2035	4 US-08-960-780-10	Sequence 10, Appl
34	17	0.5	2035	4 US-09-073-898-10	Sequence 10, Appl
35	17	0.5	2521	4 US-09-115-446-1	Sequence 1, Appl
36	17	0.5	2526	4 US-09-115-446-5	Sequence 5, Appl
37	17	0.5	4165	1 US-08-095-737-1	Sequence 1, Appl
38	17	0.5	4165	1 US-08-480-145-1	Sequence 1, Appl
39	17	0.5	4165	2 US-08-477-389-1	Sequence 1, Appl
40	17	0.5	4437	1 US-08-559-303B-72	Sequence 72, Appl
41	17	0.5	4437	4 US-09-175-828-72	Sequence 72, Appl
42	17	0.5	441529	4 US-09-103-840A-1	Sequence 1, Appl
43	16	0.5	36	1 US-08-137-117D-48	Sequence 48, Appl
44	16	0.5	36	1 US-08-436-717-48	Sequence 48, Appl
45	16	0.5	84	4 US-08-943-731-153	Sequence 153, App
46	16	0.5	132	1 US-08-479-577-1	Sequence 1, Appl
47	16	0.5	132	2 US-08-756-977-1	Sequence 1, Appl
48	16	0.5	132	5 PCT-US93-05331-5	Sequence 5, Appl
49	16	0.5	284	3 US-08-621-018B-20	Sequence 20, Appl
50	16	0.5	300	3 US-08-171-385-26	Sequence 26, Appl
51	16	0.5	311	4 US-08-361-441B-26	Sequence 26, Appl
52	16	0.5	311	4 US-08-916-576B-40	Sequence 40, Appl
53	16	0.5	355	4 US-08-991-789A-78	Sequence 78, Appl
54	16	0.5	355	4 US-08-991-789A-28	Sequence 28, App
55	16	0.5	424	1 US-08-470-179-100	Sequence 100, App
56	16	0.5	434	4 US-08-943-731-127	Sequence 127, App
57	16	0.5	520	3 US-08-899-786-14	Sequence 14, Appl
58	16	0.5	562	3 US-08-899-786-16	Sequence 16, Appl
59	16	0.5	573	4 US-08-906-156A-67	Sequence 67, Appl
60	16	0.5	574	3 US-08-899-786-15	Sequence 15, Appl
61	16	0.5	590	4 US-08-906-156A-19	Sequence 19, Appl
62	16	0.5	603	4 US-09-385-982-233	Sequence 233, App
63	16	0.5	607	4 US-09-328-111-205	Sequence 205, App
64	16	0.5	625	4 US-09-385-982-101	Sequence 101, App
65	16	0.5	697	1 US-08-468-347-25	Sequence 25, Appl
66	16	0.5	697	1 US-08-467-389-25	Sequence 25, Appl
67	16	0.5	697	2 US-08-779-379-25	Sequence 25, Appl
68	16	0.5	697	2 US-08-469-219-25	Sequence 25, Appl
69	16	0.5	697	4 US-09-228-152-25	Sequence 25, Appl
70	16	0.5	732	4 US-08-916-576B-5	Sequence 5, Appl
71	16	0.5	733	1 US-08-152-485-3	Sequence 3, Appl
72	16	0.5	733	1 US-08-463-089-3	Sequence 3, Appl
73	16	0.5	733	1 US-08-461-360A-3	Sequence 3, Appl
74	16	0.5	733	1 US-08-461-359-3	Sequence 3, Appl
75	16	0.5	733	5 PCT-US94-12904-3	Sequence 3, Appl
76	16	0.5	795	2 US-08-822-028-11	Sequence 11, Appl
77	16	0.5	795	4 US-08-479-285-11	Sequence 11, Appl
78	16	0.5	803	3 US-08-454-928-9	Sequence 9, Appl
79	16	0.5	843	4 US-09-296-284-24	Sequence 24, Appl
80	16	0.5	897	5 PCT-US96-05330A-303	Sequence 303, App
81	16	0.5	921	4 US-09-296-284-3	Sequence 3, Appl
82	16	0.5	1047	2 US-08-535-276-7	Sequence 7, Appl
83	16	0.5	1054	1 US-08-152-485-1	Sequence 1, Appl
84	16	0.5	1054	1 US-08-463-089-1	Sequence 1, Appl
85	16	0.5	1054	1 US-08-461-360A-1	Sequence 1, Appl
86	16	0.5	1054	1 US-08-461-359-1	Sequence 1, Appl
87	16	0.5	1054	5 PCT-US94-12904-1	Sequence 1, Appl
88	16	0.5	1081	1 US-08-832-883-57	Sequence 57, Appl
89	16	0.5	1081	2 US-08-832-877-57	Sequence 57, Appl
90	16	0.5	1103	3 US-08-899-786-13	Sequence 13, Appl
91	16	0.5	1104	1 US-08-823-399B-36	Sequence 36, Appl
92	16	0.5	1134	4 US-08-432-994A-9	Sequence 9, Appl
93	16	0.5	1166	5 US-08-423-399B-34	Sequence 34, Appl
94	16	0.5	1173	5 PCT-US96-12198B-1	Sequence 1, Appl
95	16	0.5	1178	2 US-08-933-750C-88	Sequence 88, Appl
96	16	0.5	1178	2 US-09-234-613-88	Sequence 88, Appl
97	16	0.5	1244	2 US-08-204-288-3	Sequence 3, Appl
98	16	0.5	1284	2 US-07-882-790-3	Sequence 3, Appl
99	16	0.5	1330	2 US-08-868-288A-4	Sequence 4, Appl
100	16	0.5			

ALIGNMENTS

RESULT 1
US-08-870-518-34
Sequence 34, Application US/08870518
Patent No. 5925566
GENERAL INFORMATION:
APPLICANT: Davis, Roger J.
TITLE OF INVENTION: Galcheva-Gargova, Zoya
TITLE OF INVENTION: NON-ACTIVATED RECEPTOR COMPLEX
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/870,518
FILING DATE: 06-JUN-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/019,219
FILING DATE: 06-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Fasse, Peter J.
REGISTRATION NUMBER: 32,983
REFERENCE/DOCKET NUMBER: 04020/102001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 1094 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-870-518-34

Query Match 0.6%; Score 18; DB 2; Length 1094;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 766 ttctcctcttgatggct 783
|||||
DB 755 ttttcccttcttgatggct 772

RESULT 2
US-08-747-221B-16/C
Sequence 16, Application US/08747221B
Patent No. 6063610
GENERAL INFORMATION:
APPLICANT: Silver, Gary W.
TITLE OF INVENTION: No. 6063610el Carboxylesterase Nucleic Acid
TITLE OF INVENTION: Molecules, Proteins and Uses Thereof
NUMBER OF SEQUENCES: 66
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Carol Talkington Verser, Ph.D.
ADDRESS: Heska Corporation
STREET: 1825 Sharp Point Drive

CITY: Fort Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: Wordperfect for Windows, Version 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/747,221B
FILING DATE: No. 6063610ember 12, 1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: FC-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 970/493-7272
TELEFAX: 970/484-9505
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 1515 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1515
FEATURE:
NAME/KEY: Asx = Asn or Asp
LOCATION: 298
US-08-747-221B-16

Query Match 0.6%; Score 18; DB 3; Length 1515;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1002 ctccaattcgaattcc 1019
|||||
DB 363 ctccaattcgaattcc 346

RESULT 3
US-08-747-221B-17
Sequence 17, Application US/08747221B
Patent No. 6063610
GENERAL INFORMATION:
APPLICANT: Silver, Gary W.
TITLE OF INVENTION: No. 6063610el Carboxylesterase Nucleic Acid
TITLE OF INVENTION: Molecules, Proteins and Uses Thereof
NUMBER OF SEQUENCES: 66
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Carol Talkington Verser, Ph.D.
ADDRESS: Heska Corporation
STREET: 1825 Sharp Point Drive
CITY: Fort Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: Wordperfect for Windows, Version 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/747,221B
FILING DATE: No. 6063610ember 12, 1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:

NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: FC-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 970/493-7272
TELEFAX: 970/484-9505
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 1515 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-747-221B-17

Query Match 0.6%; Score 18; DB 3; Length 1515;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1002 ctccaattcagaatcc 1019
DB 1153 CTCCAATTCAGAAATCC 1170

RESULT 4
US-09-005-051-16/c

Sequence 16, Application US/09005051
Patent No. 6291222

GENERAL INFORMATION:

APPLICANT: Silver, Gary W.

APPLICANT: Wisniewski, Nancy

TITLE OF INVENTION: No. 6291222e1 Carboxylesterase Nucleic Acid

NUMBER OF SEQUENCES: 66

CORRESPONDENCE ADDRESS:

ADDRESSEE: Carol Talkington Verser, Ph.D.

STREET: 1825 Sharp Point Drive

CITY: Fort Collins

STATE: Colorado

COUNTRY: USA

ZIP: 80525

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: Windows 95

SOFTWARE: Wordperfect for Windows, Version 7.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/005, 051

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/747,221

FILING DATE: No. 6291222e1, December 12, 1996

ATTORNEY/AGENT INFORMATION:

NAME: Verser, Carol Talkington

REGISTRATION NUMBER: 37,459

REFERENCE/DOCKET NUMBER: FC-1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 970/493-7272

TELEFAX: 970/484-9505

INFORMATION FOR SEQ ID NO: 16:

SEQUENCE CHARACTERISTICS:

LENGTH: 1515 nucleotides

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: CDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 1..1515

FEATURE:

NAME/KEY: Asx = Asn or Asp
LOCATION: 298
US-09-005-051-16

Query Match 0.6%; Score 18; DB 4; Length 1515;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1002 ctccaattcagaatcc 1019
DB 363 CTCCAATTCAGAAATCC 346

RESULT 5
US-09-005-051-17

Sequence 17, Application US/09005051
Patent No. 6291222

GENERAL INFORMATION:

APPLICANT: Silver, Gary W.

APPLICANT: Wisniewski, Nancy

TITLE OF INVENTION: No. 6291222e1 Carboxylesterase Nucleic Acid

NUMBER OF SEQUENCES: 66

CORRESPONDENCE ADDRESS:

ADDRESSEE: Carol Talkington Verser, Ph.D.

STREET: 1825 Sharp Point Drive

CITY: Fort Collins

STATE: Colorado

COUNTRY: USA

ZIP: 80525

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: Windows 95

SOFTWARE: Wordperfect for Windows, Version 7.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/005, 051

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/747,221

FILING DATE: No. 6291222e1, December 12, 1996

ATTORNEY/AGENT INFORMATION:

NAME: Verser, Carol Talkington

REGISTRATION NUMBER: 37,459

REFERENCE/DOCKET NUMBER: FC-1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 970/493-7272

TELEFAX: 970/484-9505

INFORMATION FOR SEQ ID NO: 17:

SEQUENCE CHARACTERISTICS:

LENGTH: 1515 nucleotides

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: CDNA

US-09-005-051-17

Query Match 0.6%; Score 18; DB 4; Length 1515;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1002 ctccaattcagaatcc 1019
DB 1153 CTCCAATTCAGAAATCC 1170

RESULT 6

US-09-005-051-17

Patent No. 5213972

APPLICANT: MCCANDLISS, RUSSELL J.; ANDERSON, DAVID M.
TITLE OF INVENTION: FERMENTATION PROCESS FOR THE PRODUCTION
OF PYRIMIDINE DEOXYRIBONUCLEOSIDES
NUMBER OF SEQUENCES: 9
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/448,158
FILING DATE: 08-DEC-1989
SEQ ID NO: 6:
LENGTH: 1611
5213972-6

Query Match 0.6%; Score 18; DB 6; Length 1611;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 343 agacttaagacttgc 360
|||||
979 AGGATTAAAGGACTTGT 962

RESULT 7

US-08-747-221B-60/c
Sequence 60, Application US/08747221B
Patent No. 6063610

GENERAL INFORMATION:

APPLICANT: Silver, Gary W.

APPLICANT: Wisniewski, Nancy

TITLE OF INVENTION: No. 6063610el Carboxylesterase Nucleic Acid

TITLE OF INVENTION: Molecules, Proteins and Uses Thereof

NUMBER OF SEQUENCES: 66

CORRESPONDENCE ADDRESS:

ADDRESSEE: Carol Talkington Verser, Ph.D.

ADDRESSEE: Heska Corporation

STREET: 1825 Sharp Point Drive

CITY: Fort Collins

STATE: Colorado

COUNTRY: USA

ZIP: 80525

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: Windows 95

SOFTWARE: Wordperfect for Windows, Version 7.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/747,221B

FILING DATE: No. 6063610ember 12, 1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Verser, Carol Talkington

REGISTRATION NUMBER: 37,459

REFERENCE/DOCKET NUMBER: FC-1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 970/493-7272

TELEFAX: 970/484-9505

INFORMATION FOR SEQ ID NO: 60:

SEQUENCE CHARACTERISTICS:

LENGTH: 1650 nucleotides

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: CDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 1..1650

FEATURE:

NAME/KEY: ASX - Asn or Asp

LOCATION: 433

US-08-747-221B-60

Query Match 0.6%; Score 18; DB 3; Length 1650;
Best Local Similarity 100.0%; Pred. No. 47;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1002 ctcacaattcagaatcc 1019
|||||
DB 438 CTCCAATTCAAGAAATCC 481

RESULT 8

US-08-747-221B-61
Sequence 61, Application US/08747221B
Patent No. 6063610

GENERAL INFORMATION:

APPLICANT: Silver, Gary W.

APPLICANT: Wisniewski, Nancy

TITLE OF INVENTION: No. 6063610el Carboxylesterase Nucleic Acid

TITLE OF INVENTION: Molecules, Proteins and Uses Thereof

NUMBER OF SEQUENCES: 66

CORRESPONDENCE ADDRESS:

ADDRESSEE: Carol Talkington Verser, Ph.D.

ADDRESSEE: Heska Corporation

STREET: 1825 Sharp Point Drive

CITY: Fort Collins

STATE: Colorado

COUNTRY: USA

ZIP: 80525

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: Windows 95

SOFTWARE: Wordperfect for Windows, Version 7.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/747,221B

FILING DATE: No. 6063610ember 12, 1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Verser, Carol Talkington

REGISTRATION NUMBER: 37,459

REFERENCE/DOCKET NUMBER: FC-1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 970/493-7272

TELEFAX: 970/484-9505

INFORMATION FOR SEQ ID NO: 61:

SEQUENCE CHARACTERISTICS:

LENGTH: 1650 nucleotides

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: CDNA

US-08-747-221B-61

Query Match 0.6%; Score 18; DB 3; Length 1650;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1002 ctcacaattcagaatcc 1019
|||||
DB 1153 CTCCAATTCAAGAAATCC 1170

RESULT 9

US-09-005-051-60/c
Sequence 60, Application US/09005051
Patent No. 6291222

GENERAL INFORMATION:

APPLICANT: Silver, Gary W.

APPLICANT: Wisniewski, Nancy

TITLE OF INVENTION: No. 6291222el Carboxylesterase Nucleic Acid

TITLE OF INVENTION: Molecules, Proteins and Uses Thereof

NUMBER OF SEQUENCES: 66

CORRESPONDENCE ADDRESS:

ADDRESSEE: Carol Talkington Verser, Ph.D.

ADDRESSEE: Heska Corporation

STREET: 1825 Sharp Point Drive
CITY: Fort Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: Wordperfect for Windows, Version 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/005,051
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/747,221
FILING DATE: No. 6291222ember 12, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: FC-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 970/493-7272
TELEFAX: 970/484-9505
INFORMATION FOR SEQ ID NO: 60:
SEQUENCE CHARACTERISTICS:
LENGTH: 1650 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1650
FEATURE: Asx - Asn or Asp
NAME/KEY: 433
LOCATION:
US-09-005-051-60

Query Match 0.6%; Score 18; DB 4; Length 1650;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1002 ctccaattcagaatcc 1019
Db 498 CTCCAATTCAGAAATCC 481

RESULT 10
US-09-005-051-61
Sequence 61, Application US/09005051
Patent No. 6291222
GENERAL INFORMATION:
APPLICANT: Silver, Gary W.
TITLE OF INVENTION: No. 6291222el Carboxylesterase Nucleic Acid
NUMBER OF INVENTION: Molecules, Proteins and Uses Thereof
NUMBER OF SEQUENCES: 66
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carol Talkington Verser, Ph.D.
ADDRESSEE: Heska Corporation
STREET: 1825 Sharp Point Drive
CITY: Fort Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: Wordperfect for Windows, Version 7.0
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/005,051
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/747,221
FILING DATE: No. 6291222ember 12, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: FC-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 970/493-7272
TELEFAX: 970/484-9505
INFORMATION FOR SEQ ID NO: 61:
SEQUENCE CHARACTERISTICS:
LENGTH: 1650 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-09-005-051-61

Query Match 0.6%; Score 18; DB 4; Length 1650;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1002 ctccaattcagaatcc 1019
Db 1153 CTCCAATTCAGAAATCC 1170

RESULT 11
US-08-857-213-2/c
Sequence 2, Application US/08857213
Patent No. 6054290
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Murry, Lynn E.
TITLE OF INVENTION: HUMAN VESICLE BINDING PROTEIN
NUMBER OF INVENTION: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/857,213
FILING DATE: Herewith
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0297 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1721 base pairs
TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: FIBRNGT01
CLONE: 148415
US-08-857-213-2

Query Match 0.6%; Score 18; DB 3; Length 1721;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 36 atgaattcataaggac 53
|||||
Db 1538 ATGAATTCATTAAGAC 1521

USUT 12
-08-139-937-11
Sequence 11, Application US/08139937
Patent No. 5821070

GENERAL INFORMATION:

APPLICANT: LEE, WEN-HWA
TITLE OF INVENTION: CELLULAR GENES ENCODING
TITLE OF INVENTION: RETINOBLASTOMA-ASSOCIATED PROTEINS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: CAMPBELL AND FLORES
STREET: 4370 LA JOLLA VILLAGE DRIVE
CITY: SAN DIEGO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 92122

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/139,937
FILING DATE: 20-OCT-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/979,156
FILING DATE: 20-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: CAMPBELL, CATRYN
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-CJ 9370
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-535-9001
TELEFAX: 619-535-8949
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1800 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-139-937-11

Query Match 0.6%; Score 18; DB 1; Length 1800;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2831 aaagaacctcgtaaaa 2848
|||||
Db 307 AAAGAACCCTCGAAAAA 324

RESULT 13.

PCT-US93-11310-11
Sequence 11, Application PC/TUS9311310

GENERAL INFORMATION:

APPLICANT: BOARD OF REGENTS OF THE UNIVERSITY OF TEXAS SYSTEM
TITLE OF INVENTION: CELLULAR GENES ENCODING
TITLE OF INVENTION: RETINOBLASTOMA-ASSOCIATED PROTEINS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: CAMPBELL AND FLORES
STREET: 4370 LA JOLLA VILLAGE DRIVE
CITY: SAN DIEGO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 92122

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/11310
FILING DATE: 19-NOV-1993
CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: CAMPBELL, CATRYN
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: FP-CJ 9790
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-535-9001
TELEFAX: 619-535-8949
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1800 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
PCT-US93-11310-11

Query Match 0.6%; Score 18; DB 5; Length 1800;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2831 aaagaacctcgtaaaa 2848
|||||
Db 307 AAAGAACCCTCGAAAAA 324

RESULT 14

US-08-747-221B-13/C
Sequence 13, Application US/08747221B

Patent No. 6063610

GENERAL INFORMATION:

APPLICANT: Silver, Gary W.
TITLE OF INVENTION: No. 6063610e1 Carboxylesterase Nucleic Acid
TITLE OF INVENTION: Molecules, Proteins and Uses Thereof
NUMBER OF SEQUENCES: 66
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carol Talkington Verser, Ph.D.
STREET: 1825 Sharp Point Drive
CITY: Fort Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: Wordperfect for Windows, Version 7.0
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/747,221B
FILING DATE: No. 6063610ember 12, 1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: FC-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 970/493-7272
TELEFAX: 970/484-9505
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 1982 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 31..1517
FEATURE:
NAME/KEY: Asx = Asn or Asp
LOCATION: 300
US-08-747-221B-13

Query Match 0.6%; Score 18; DB 3; Length 1982;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1002 ctccaattcagaatcc 1019
|||||

DB 365 CTCGAATTCAGAAATCC 348

RESULT 15
US-08-747-221B-15
Sequence 15, Application US/08747221B
Patent No. 6063610
GENERAL INFORMATION:
APPLICANT: Silver, Gary W.
TITLE OF INVENTION: No. 6063610el Carboxylesterase Nucleic Acid
TITLE OF INVENTION: Molecules, Proteins and Uses Thereof
NUMBER OF SEQUENCES: 66
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carol Talkington Verser, Ph.D.
ADDRESSEE: Heska Corporation
STREET: 1825 Sharp Point Drive
CITY: Fort Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: WordPerfect for Windows, Version 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/747,221B
FILING DATE: No. 6063610ember 12, 1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: FC-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 970/493-7272
TELEFAX: 970/484-9505
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 1982 nucleotides
TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-747-221B-15

Query Match 0.6%; Score 18; DB 3; Length 1982;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1002 ctccaattcagaatcc 1019
|||||

DB 1618 CTCGAATTCAGAAATCC 1635

RESULT 16
US-09-005-051-13/C
Sequence 13, Application US/09005051
Patent No. 6291222
GENERAL INFORMATION:
APPLICANT: Silver, Gary W.
TITLE OF INVENTION: No. 6291222el Carboxylesterase Nucleic Acid
TITLE OF INVENTION: Molecules, Proteins and Uses Thereof
NUMBER OF SEQUENCES: 66
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carol Talkington Verser, Ph.D.
ADDRESSEE: Heska Corporation
STREET: 1825 Sharp Point Drive
CITY: Fort Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: WordPerfect for Windows, Version 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/005,051
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/747,221
FILING DATE: No. 6291222ember 12, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: FC-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 970/493-7272
TELEFAX: 970/484-9505
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 1982 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 31..1517
FEATURE:
NAME/KEY: Asx = Asn or Asp
LOCATION: 300
US-09-005-051-13

Query Match 0.6%; Score 18; DB 4; Length 1982;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1002 ctccaattcagaatcc 1019

Db 365 CTCCAATTGAGAAATCC 348

RESULT 17

US-09-005-051-15

Sequence 15, Application US/09005051

Patent No. 6291222

GENERAL INFORMATION:

APPLICANT: Silver, Gary W.

APPLICANT: Wisniewski, Nancy

TITLE OF INVENTION: No. 6291222el Carboxylesterase Nucleic Acid

TITLE OF INVENTION: Molecules, Proteins and Uses Thereof

NUMBER OF SEQUENCES: 66

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Carol Talkington Verser, Ph.D.

ADDRESSEE: Heska Corporation

STREET: 1825 Sharp Point Drive

CITY: Fort Collins

STATE: Colorado

COUNTRY: USA

ZIP: 80525

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: Windows 95

SOFTWARE: Wordperfect for Windows, Version 7.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/005,051

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/747,221

FILING DATE: No. 6291222el December 12, 1996

ATTORNEY/AGENT INFORMATION:

NAME: Verser, Carol Talkington

REGISTRATION NUMBER: 37,459

REFERENCE/DOCKET NUMBER: FC-1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 970/493-7272

TELEFAX: 970/484-9505

INFORMATION FOR SEQ ID NO: 15:

SEQUENCE CHARACTERISTICS:

LENGTH: 1982 nucleotides

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: CDNA

-09-005-051-15

Query Match 0.6%; Score 18; DB 4; Length 1982;

Best Local Similarity 100.0%; Pred. No. 47;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1002 ctccaattcagaatcc 1019

Db 1618 CTCCAATTGAGAAATCC 1635

US-08-747-221B-57/c

Sequence 57, Application US/08747221B

Patent No. 6063610

GENERAL INFORMATION:

APPLICANT: Silver, Gary W.

APPLICANT: Wisniewski, Nancy

TITLE OF INVENTION: No. 6063610el Carboxylesterase Nucleic Acid

TITLE OF INVENTION: Molecules, Proteins and Uses Thereof

NUMBER OF SEQUENCES: 66

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Carol Talkington Verser, Ph.D.

ADDRESSEE: Heska Corporation

STREET: 1825 Sharp Point Drive

CITY: Fort Collins

STATE: Colorado

COUNTRY: USA

ZIP: 80525

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: Windows 95

SOFTWARE: Wordperfect for Windows, Version 7.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/747,221B

FILING DATE: No. 6063610el December 12, 1996

CLASSIFICATION: 435

Query Match 0.6%; Score 18; DB 3; Length 2144;

Best Local Similarity 100.0%; Pred. No. 47;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1002 ctccaattcagaatcc 1019

Db 527 CTCCAATTGAGAAATCC 510

US-08-747-221B-59

Sequence 59, Application US/08747221B

Patent No. 6063610

GENERAL INFORMATION:

APPLICANT: Silver, Gary W.

APPLICANT: Wisniewski, Nancy

TITLE OF INVENTION: No. 6063610el Carboxylesterase Nucleic Acid

TITLE OF INVENTION: Molecules, Proteins and Uses Thereof

NUMBER OF SEQUENCES: 66

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Carol Talkington Verser, Ph.D.

ADDRESSEE: Heska Corporation

STREET: 1825 Sharp Point Drive

CITY: Fort Collins

STATE: Colorado

COUNTRY: USA

ZIP: 80525

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: Windows 95

SOFTWARE: Wordperfect for Windows, Version 7.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/747,221B

FILING DATE: No. 6063610el December 12, 1996

CLASSIFICATION: 435

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Verser, Carol Talkington
; REGISTRATION NUMBER: 37,459
; REFERENCE/DOCKET NUMBER: FC-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 970/493-7272
; TELEFAX: 970/484-9505
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2144 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-747-221B-59

Query Match 0.6%; Score 18; DB 3; Length 2144;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1002 ctccaattcagaatcc 1019
1618 CTCCAATTGAGAAATCC 1635

RESULT 20
US-09-005-051-57/C
; Sequence 57, Application US/09005051
; Patent No. 6291222
; GENERAL INFORMATION:
; APPLICANT: Silver, Gary W.
; TITLE OF INVENTION: No. 6291222el Carboxylesterase Nucleic Acid
; TITLE OF INVENTION: Molecules, Proteins and Uses Thereof
; NUMBER OF SEQUENCES: 66
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Carol Talkington Verser, Ph.D.
; STREET: 1825 Sharp Point Drive
; CITY: Fort Collins
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80525
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: Wordperfect for Windows, Version 7.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/005,051
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/747,221
; FILING DATE: No. 6291222ember 12, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Verser, Carol Talkington
; REGISTRATION NUMBER: 37,459
; REFERENCE/DOCKET NUMBER: FC-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 970/493-7272
; TELEFAX: 970/484-9505
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2144 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 30..1682
```

```

; FEATURE:
; NAME/KEY: Asx - Asn or Asp
; LOCATION: 462
; US-09-005-051-57

Query Match 0.6%; Score 18; DB 4; Length 2144;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1002 ctccaattcagaatcc 1019
Db 527 CTCCAATTGAGAAATCC 510

RESULT 21
US-09-005-051-59
; Sequence 59, Application US/09005051
; Patent No. 6291222
; GENERAL INFORMATION:
; APPLICANT: Silver, Gary W.
; TITLE OF INVENTION: No. 6291222el Carboxylesterase Nucleic Acid
; TITLE OF INVENTION: Molecules, Proteins and Uses Thereof
; NUMBER OF SEQUENCES: 66
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Carol Talkington Verser, Ph.D.
; STREET: 1825 Sharp Point Drive
; CITY: Fort Collins
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80525
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: Wordperfect for Windows, Version 7.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/005,051
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/747,221
; FILING DATE: No. 6291222ember 12, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Verser, Carol Talkington
; REGISTRATION NUMBER: 37,459
; REFERENCE/DOCKET NUMBER: FC-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 970/493-7272
; TELEFAX: 970/484-9505
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2144 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-09-005-051-59

Query Match 0.6%; Score 18; DB 4; Length 2144;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1002 ctccaattcagaatcc 1019
Db 1618 CTCCAATTGAGAAATCC 1635

RESULT 22
US-09-192-104-1
```

```
Sequence 1, Application US/09192104B
Patent No. 6184020
GENERAL INFORMATION:
APPLICANT: Alexander Blinkovsky
APPLICANT: Tony Byun
APPLICANT: Alan V. Klotz
APPLICANT: Alan Sloma
APPLICANT: Maria Tang
APPLICANT: Mikiso Fujii
APPLICANT: Chigusa Marumoto
APPLICANT: Lene Venke Kofod
TITLE OF INVENTION: Polypeptides Having Amino-peptidase
FILE REFERENCE: 5379.200-US
CURRENT APPLICATION NUMBER: US/09/192.104B
EARLIER FILING DATE: 1998-11-13
EARLIER APPLICATION NUMBER: 60/069719
EARLIER FILING DATE: 1997-12-16
EARLIER APPLICATION NUMBER: 1465/97
EARLIER FILING DATE: 1997-12-16
EARLIER APPLICATION NUMBER: PA 1998 00670
EARLIER FILING DATE: 1998-05-15
NUMBER OF SEQ ID NOS: 9
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 3000
TYPE: DNA
ORGANISM: Sphingomonas
US-09-192-104-1
```

```
Query Match 0.6%; Score 18; DB 4; Length 3000;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2127 gctgtgtgctgtgaagc 2144
|||||
Db 1911 gctgtgtgctgtgaagc 1928
```

```
RESULT 23
US-09-543-446-1
Sequence 1, Application US/09543446
Patent No. 6303360
GENERAL INFORMATION:
APPLICANT: Alexander Blinkovsky
APPLICANT: Tony Byun
APPLICANT: Alan V. Klotz
APPLICANT: Alan Sloma
APPLICANT: Maria Tang
APPLICANT: Mikiso Fujii
APPLICANT: Chigusa Marumoto
APPLICANT: Lene Venke Kofod
TITLE OF INVENTION: Polypeptides Having Amino-peptidase
FILE REFERENCE: 5379.210-US
CURRENT APPLICATION NUMBER: US/09/543.446
EARLIER FILING DATE: 2000-04-05
EARLIER APPLICATION NUMBER: 60/069719
EARLIER FILING DATE: 1997-12-16
EARLIER APPLICATION NUMBER: 1465/97
EARLIER FILING DATE: 1997-12-16
EARLIER APPLICATION NUMBER: PA 1998 00670
EARLIER FILING DATE: 1998-05-15
EARLIER APPLICATION NUMBER: 09/192.104
EARLIER FILING DATE: 1998-11-13
NUMBER OF SEQ ID NOS: 9
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 3000
TYPE: DNA
ORGANISM: Sphingomonas
US-09-543-446-1
```

```
Query Match 0.6%; Score 18; DB 4; Length 3000;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2127 gctgtgtgctgtgaagc 2144
|||||
Db 1911 gctgtgtgctgtgaagc 1928
```

```
RESULT 24
US-08-699-103B-1/c
Sequence 1, Application US/08699103B
Patent No. 6107462
GENERAL INFORMATION:
APPLICANT: Rine, Jasper D.
APPLICANT: Hampton, Randolph
TITLE OF INVENTION: GENES AND PROTEINS CONTROLLING
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESS: Fish & Richardson P.C.
STREET: 2200 Sand Hill Road, Suite 100
CITY: Menlo Park
STATE: CA
COUNTRY: USA
ZIP: 94025
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/699.103B
FILING DATE: 16-AUG-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/002,581
FILING DATE: 17-AUG-1995
ATTORNEY/AGENT INFORMATION:
NAME: Green, Grant D.
REGISTRATION NUMBER: 31,259
REFERENCE/DOCKET NUMBER: 09272/005001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/322-5070
TELEFAX: 650/854-0875
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4982 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-08-699-103B-1
```

```
Query Match 0.6%; Score 18; DB 3; Length 4982;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 391 cccttgtgaagatattg 408
|||||
Db 1743 CCCTTGTGAAGATATTG 1726
```

```
RESULT 25
US-09-037-990B-27
Sequence 27, Application US/09037990B
Patent No. 6248519
GENERAL INFORMATION:
APPLICANT: ENGEL, Stacia R.
DESCENZO, Richard A.
MORENZONI, Richard A.
US-09-037-990B-27
```

```

1 IRELAN, Nancy A.
2 TITLE OF INVENTION: DETECTION OF FERMENTATION-RELATED
3 MICROORGANISMS
4 NUMBER OF SEQUENCES: 100
5 CORRESPONDENCE ADDRESSES:
6 ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
7 STREET: 6300 Sears Tower, 233 South Wacker Drive
8 CITY: Chicago
9 STATE: Illinois
10 COUNTRY: United States of America
11 ZIP: 60606-6402
12 COMPUTER READABLE FORM:
13 MEDIUM TYPE: Floppy disk
14 OPERATING SYSTEM: PC-DOS/MS-DOS
15 SOFTWARE: Patentin Release #1.0, Version #1.30
16 CURRENT APPLICATION DATA:
17 APPLICATION NUMBER: US/09/037,990B
18 FILING DATE: 11-Mar-1999
19 CLASSIFICATION: <Unknown>
20 PRIOR APPLICATION DATA:
21 APPLICATION NUMBER: <Unknown>
22 FILING DATE: <Unknown>
23 ATTORNEY/AGENT INFORMATION:
24 NAME: Sharp, Jeffrey S.
25 REGISTRATION NUMBER: 31,879
26 REFERENCE/DOCKET NUMBER: 29520/30001
27 TELECOMMUNICATION INFORMATION:
28 TELEPHONE: 312/474-6300
29 TELEFAX: 312/474-0448
30 TELEX: <Unknown>
31 INFORMATION FOR SEQ ID NO: 27:
32 SEQUENCE CHARACTERISTICS:
33 LENGTH: 290 base pairs
34 TYPE: nucleic acid
35 STRANDEDNESS: single
36 TOPOLOGY: linear
37 MOLECULE TYPE: DNA (genomic)
38 SEQUENCE DESCRIPTION: SEQ ID NO: 27:
39 US-09-037-990B-27
40
41 Query Match          0.5%; Score 17; DB 4; Length 290;
42 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
43 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
44
45 Oy      1152 aaacaacaaacaaca 1168
46         |||||||
47 Db      209 AAAAACAACAAACA 225
48
49 RESULT 26
50 US-08-991-789A-171/c
51 Sequence 171, Application US/08991789A
52 Patent No. 6225054
53 GENERAL INFORMATION:
54 APPLICANT: Frudakis, Tony N.
55             Smith, John M.
56             Reed, Steven G.
57 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
58 TREATMENT AND DIAGNOSIS OF BREAST CANCER
59 NUMBER OF SEQUENCES: 292
60 CORRESPONDENCE ADDRESS:
61 ADDRESSEE: Seed IP Law Group
62 STREET: 701 Fifth Avenue, Suite 6300
63 CITY: Seattle
64 STATE: Washington
65 COUNTRY: USA
66 ZIP: 98104-7092
67 COMPUTER READABLE FORM:
68 MEDIUM TYPE: Floppy disk
69 COMPUTER: IBM PC compatible
70 OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

1      SOFTWARE: Patentln Release #1.0, Version #1.30
2      CURRENT APPLICATION DATA:
3      APPLICATION NUMBER: US/08/991,789A
4      FILING DATE: 11-Dec-1997
5      CLASSIFICATION: <Unknown>
6      ATTORNEY/AGENT INFORMATION:
7      NAME: Potler, Jane E. R.
8      REGISTRATION NUMBER: 33,332
9      REFERENCE/DOCKET NUMBER: 210121.419C3
10     TELECOMMUNICATION INFORMATION:
11     TELEPHONE: (206) 622-4900
12     TELEFAX: (206) 682-6031
13     INFORMATION FOR SEQ ID NO: 171:
14     SEQUENCE CHARACTERISTICS:
15     LENGTH: 383 base pairs
16     TYPE: nucleic acid
17     STRANDEDNESS: single
18     TOPOLOGY: linear
19     SEQUENCE DESCRIPTION: SEQ ID NO: 171:
20     US-08-991-789A-171
21
22     Query Match          0.5%; Score 17; DB 4; Length 383;
23     Best Local Similarity 100.0%; Pred. No. 14e+02;
24     Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
25
26     QY 409 acatactatttactctg 425
27     |||||
28     Db 35 ACATTATTTTTAACTTG 19
29
30 RESULT 27
31 US-09-037-990B-8
32 : Sequence 8, Application US/09037990B
33 : Patent No. 6248519
34 : GENERAL INFORMATION:
35 : APPLICANT: ENGEL, Stacia R.
36 : DESCENZO, Richard A.
37 : MORENZONI, Richard A.
38 : IRELAN, Nancy A.
39 : TITLE OF INVENTION: DETECTION OF FERMENTATION-RELATED
40 : MICROORGANISMS
41 : NUMBER OF SEQUENCES: 100
42 : CORRESPONDENCE ADDRESS:
43 : ADDRESSEE: Marshall, O'Toole, Gersteln, Murray & Borun
44 : STREET: 6300 Sears Tower, 233 South Wacker Drive
45 : CITY: Chicago
46 : STATE: Illinois
47 : COUNTRY: United States of America
48 : ZIP: 60606-6402
49 : COMPUTER READABLE FORM:
50 : MEDIUM TYPE: floppy disk
51 : COMPUTER: IBM PC compatible
52 : OPERATING SYSTEM: PC-DOS/MS-DOS
53 : SOFTWARE: Patentln Release #1.0, Version #1.30
54 : CURRENT APPLICATION DATA:
55 : APPLICATION NUMBER: US/09/037,990B
56 : FILING DATE: 11-Mar-1999
57 : CLASSIFICATION: <Unknown>
58 : PRIOR APPLICATION DATA:
59 : APPLICATION NUMBER: <Unknown>
60 : FILING DATE: <Unknown>
61 : ATTORNEY/AGENT INFORMATION:
62 : NAME: Sharp, Jeffrey S.
63 : REGISTRATION NUMBER: 31,879
64 : REFERENCE/DOCKET NUMBER: 29520/30001
65 : TELECOMMUNICATION INFORMATION:
66 : TELEPHONE: 312/474-6300
67 : TELEFAX: 312/474-0448
68 : TELE: <Unknown>
69 : INFORMATION FOR SEQ ID NO: 8:
70 : SEQUENCE CHARACTERISTICS:
71 : LENGTH: 661 base pairs

```

```

?      TYPE: nucleic acid
?      STRANDEDNESS: single
?      TOPOLOGY: linear
?      MOLECULE TYPE: DNA (genomic)
?      SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-037-990B-8

```

```

Query Match      0.5%; Score 17; DB 4; Length 661;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy  1152 aaaaacaaacacacaaca 1168
      |||||
Db   209 AAAAACAAACACAAACA 225

```

RESULT 28
-08-998-416-818
Sequence 818, Application US/08998416
Patent No. 6239264

1 APPLICANT: Philippsen, Peter
 2 APPLICANT: Pohlmann, Rainer
 3 APPLICANT: Steiner, Sabine
 4 APPLICANT: Mohr, Christine
 5 APPLICANT: Wendland, Jurgen
 6 APPLICANT: Knechtel, Philipp
 7 APPLICANT: Redischung, Corinne
 8 TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSEYPTII
 9 TITLE OF INVENTION: AND USES THEREOF
 10 NUMBER OF SEQUENCES: 1152
 11 CORRESPONDENCE ADDRESSES:

```

: COMPUTER READABLE FORM:
:
: MEDIUM TYPE: Floppy disk
:
: COMPUTER: IBM PC compatible
:
: OPERATING SYSTEM: PC-DOS/MS-DOS
:
: SOFTWARE: PatentIn Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:

```

ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976/
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 818:
SEQUENCE CHARACTERISTICS:
LENGTH: 685 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAGI520UP
IS-08-998-416-818

```

Query Match Similarity      0.54; Score 17; DB 4; length 685;
Best Local Similarity      100.08; Pred. No. 14+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY	495	ctgaatgacctctctgca	511
Db	111	CTGATGGCCTCTGCA	127

RESULT 29
US-09-318-443-5/c
; Sequence 5, Application US/09318443
; Patent No. 6197947
; GENERAL INFORMATION:

```

?  APPLICANT: Hemmati-Briavanlou, Ali
?  APPLICANT: Weinstein, Daniel C.
?  TITLE OF INVENTION: TRANSLATION INITIATION FACTOR 4A111, AND METHODS OF USE
?  TITLE OF INVENTION: THEREOF
?  FILE REFERENCE: 600-1-211 N
?  CURRENT APPLICATION NUMBER: US/09/318,443
?  CURRENT FILING DATE: 1999-05-25
?  NUMBER OF SEQ ID NOS: 12
?  SOFTWARE: PatentIn Ver. 2.0
?  SEQ ID NO 5
?  LENGTH: 1536
?  TYPE: DNA
?  ORGANISM: Homo sapiens
?  US-09-318-443-5

```

```

Query Match      0.5%;  Score 17;  DB 4;  Length 1536;
Best Local Similarity 100.0%;  Pred. No. 1.4e+02;
Matches 17;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0

```

Oy	1374	caggaggtacacagcagg	1390
Db	1360	CAGGAGTACACAGCAGG	1344

RESULT 30
US-08-743-637B-183/c
; Sequence 183, Application US/08743637B
; Patent No. 5994066
; GENERAL INFORMATION:

APPLICANT: BERGERON, Michel G.
 APPLICANT: PICARD, Francois J.
 APPLICANT: OUELLETTE, Marc
 APPLICANT: ROY, Paul H.
 TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA
 TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
 TITLE OF INVENTION: IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
 TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ..
 NUMBER OF SEQUENCES: 273
 CORRESPONDENCE ADDRESS:

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/743,637B
FILING DATE: 04-NOV-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/526,840
FILING DATE: 11-SEP-1995
ATTORNEY/AGENT INFORMATION:
NAME: BAKER, Jean C.
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 850586, 90012

```
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5000
; TELEFAX: (414) 277-5591
; INFORMATION FOR SEQ ID NO: 183:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1563 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Klebsiella pneumoniae
; US-08-743-637B-183
```

```
Query Match 0.5%: Score 17; DB 2; Length 1563;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 505 tccgcagcgtctcag 521
|||||
74 TCCGCGAGGCTCTTCAG 58
```

```
RESULT 31
US-09-318-443-7/C
; Sequence 7, Application US/09318443
; Patent No. 6197947
; GENERAL INFORMATION:
; APPLICANT: Hemmati-Briyandlou, Ali
; APPLICANT: Weinstein, Daniel C.
; TITLE OF INVENTION: TRANSLATION INITIATION FACTOR 4AIII, AND METHODS OF USE
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 600-1-211 N
; CURRENT APPLICATION NUMBER: US/09/318,443
; CURRENT FILING DATE: 1999-05-25
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 1682
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-318-443-7
```

```
Query Match 0.5%: Score 17; DB 4; Length 1682;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
1374 caggagtagacacagcag 1390
|||||
1501 CAGGAGTAGACAGCAGCAG 1485
```

```
RESULT 32
US-09-178-252-26
; Sequence 26, Application US/09178252
; Patent No. 6218188
; GENERAL INFORMATION:
; APPLICANT: Cardineau, Guy A.
; APPLICANT: Steiman, Steven J.
; APPLICANT: Naraya, Kenneth E.
; TITLE OF INVENTION: Plant-Optimized Genes Encoding Pesticidal Toxins
; FILE REFERENCE: MA-714XC2
; CURRENT APPLICATION NUMBER: US/09/178,252
; CURRENT FILING DATE: 1998-10-23
; EARLIER APPLICATION NUMBER: 60/065,215
; EARLIER FILING DATE: 1997-11-12
; EARLIER APPLICATION NUMBER: 60/076,445
; EARLIER FILING DATE: 1998-03-02
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 26
```

```
; LENGTH: 1965
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic B.t. toxin gene
; US-09-178-252-26
```

```
Query Match 0.5%: Score 17; DB 4; Length 1965;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 463 accacacacacaccgc 479
|||||
Db 1519 accacacacacaccgc 1535
```

```
RESULT 33
US-08-960-780-10
; Sequence 10, Application US/08960780
; Patent No. 6204435
; GENERAL INFORMATION:
; APPLICANT: Feltelson, Gerald S.
; APPLICANT: Schieff, H. Ernest
; APPLICANT: Naraya, Kenneth E.
; APPLICANT: Stockhoff, Brian A.
; APPLICANT: Schmeltz, James
; APPLICANT: Loewer, David
; APPLICANT: Dullum, Charles Joseph
; APPLICANT: Muller-Cohn, Judy
; TITLE OF INVENTION: Stamp, Lisa
; TITLE OF INVENTION: No. 6204435el Pesticidal Toxins and Nucleotide
; TITLE OF INVENTION: Sequences Which Encode These Toxins
; NUMBER OF SEQUENCES: 134
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: US
; ZIP: 32606-6669
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/960,780
; FILING DATE: 30-OCT-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/029,848
; FILING DATE: 30-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: MA-708
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 352-375-8100
; TELEFAX: 352-372-5800
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2035 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: 158C2-PTL
; US-08-960-780-10
```

```
Query Match 0.5%: Score 17; DB 4; Length 2035;
```

Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2620 aagatggaatgaagct 2636
|||||
DB 1973 AAGATGGAATGAAGCT 1989

RESULT 34

US-09-073-898-10
; Sequence 10, Application US/09073898
; Patent No. 6242669

GENERAL INFORMATION:

APPLICANT: Fetteison, Jerald S.
APPLICANT: Schepf, H. Ernest
APPLICANT: Narva, Kenneth E.
APPLICANT: Stockhoff, Brian A.
APPLICANT: Schmeits, James
APPLICANT: Loewer, David
APPLICANT: Dullum, Charles Joseph
APPLICANT: Muller-Cohn, Judy
APPLICANT: Stamp, Lisa
APPLICANT: Morill, George
APPLICANT: Finstad-Lee, Stacey
TITLE OF INVENTION: No. 6242669e1 Pesticidal Toxins and Nucleotide
TITLE OF INVENTION: Sequences Which Encode These Toxins
NUMBER OF SEQUENCES: 144
CORRESPONDENCE ADDRESS:
ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: US
ZIP: 32606-6669

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/073.898
FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/029,848
FILING DATE: 30-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/960,780
FILING DATE: 30-OCT-1997
ATTORNEY/AGENT INFORMATION:

NAME: Sanders, Jay M.
REGISTRATION NUMBER: 39,355
REFERENCE/DOCKET NUMBER: MA-708C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 352-375-8100
TELEFAX: 352-372-5800

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:
LENGTH: 2035 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: 158C2-PT1
US-09-073-898-10

Query Match 0.5%; Score 17; DB 4; Length 2035;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2620 aagatggaatgaagct 2636

|||||
DB 1973 AAGATGGAATGAAGCT 1989

RESULT 35

US-09-115-446-1/C
; Sequence 1, Application US/09115446
; Patent No. 6165719

GENERAL INFORMATION:

APPLICANT: Chandry, George K.
APPLICANT: Gargus, Jay J.
APPLICANT: Gutman, George
APPLICANT: Fantino, Emmanuelle
APPLICANT: Kalman, Katarin
TITLE OF INVENTION: hKCA3/KCNN3 SMALL CONDUCTANCE CALCIUM
TITLE OF INVENTION: ACTIVATED POTASSIUM CHANNEL: A DIAGNOSTIC
TITLE OF INVENTION: MARKER AND THERAPEUTIC TARGET
FILE REFERENCE: 07306/014001
CURRENT APPLICATION NUMBER: US/09/115.446
CURRENT FILING DATE: 1998-07-14
EARLIER APPLICATION NUMBER: 60/052.556
EARLIER FILING DATE: 1997-07-15
EARLIER APPLICATION NUMBER: 60/070.741
EARLIER FILING DATE: 1998-01-08
NUMBER OF SEQ ID NOS: 15
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 2521
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (287)...(2479)
US-09-115-446-1

Query Match 0.5%; Score 17; DB 4; Length 2521;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 739 tgcgtcgaagctgcgg 755
|||||
DB 492 TcGTCGTAAGCTGCGG 476

RESULT 36

US-09-115-446-5/C
; Sequence 5, Application US/09115446
; Patent No. 6165719

GENERAL INFORMATION:

APPLICANT: Chandry, George K.
APPLICANT: Gargus, Jay J.
APPLICANT: Gutman, George
APPLICANT: Fantino, Emmanuelle
APPLICANT: Kalman, Katarin
TITLE OF INVENTION: hKCA3/KCNN3 SMALL CONDUCTANCE CALCIUM
TITLE OF INVENTION: ACTIVATED POTASSIUM CHANNEL: A DIAGNOSTIC
TITLE OF INVENTION: MARKER AND THERAPEUTIC TARGET
FILE REFERENCE: 07306/014001
CURRENT APPLICATION NUMBER: US/09/115.446
CURRENT FILING DATE: 1998-07-14
EARLIER APPLICATION NUMBER: 60/052.556
EARLIER FILING DATE: 1997-07-15
EARLIER APPLICATION NUMBER: 60/070.741
EARLIER FILING DATE: 1998-01-08
NUMBER OF SEQ ID NOS: 15
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5
LENGTH: 2526
TYPE: DNA
ORGANISM: Homo sapiens
US-09-115-446-5

Query Match 0.5%; Score 17; DB 4; Length 2526;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 739 tgcctcgaagctgcg 755
|||||
Db 497 TGCCTGAGAGCTGCCG 481

RESULT 37

US-08-095-737-1/C
; Sequence 1, Application US/08095737
; Patent No. 5487979
; GENERAL INFORMATION:
; APPLICANT: Diflore, Pier P
; APPLICANT: Fazioli, Francesca
; TITLE OF INVENTION: A Substrate for the Epidermal Growth
; TITLE OF INVENTION: Factor Receptor Kinase
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: United States of America
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/095,737
; FILING DATE: 19930722
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelson, Ned A
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH060.001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4165 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOETHERICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 21..2709
; US-08-095-737-1

Query Match 0.5%; Score 17; DB 1; Length 4165;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2811 atttagtgcatcttt 2827
|||||
Db 4070 ATTTAGTGCAATTTT 4054

RESULT 38

US-08-480-145-1/C
; Sequence 1, Application US/08480145
; Patent No. 5717067
; GENERAL INFORMATION:
; APPLICANT: Diflore, Pier P

APPLICANT: Fazioli, Francesca
; TITLE OF INVENTION: A Substrate for the Epidermal Growth
; TITLE OF INVENTION: Factor Receptor Kinase
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: United States of America
; ZIP: 92660

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,145
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/095,737
; FILING DATE: 22-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelson, Ned A
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH060.001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4165 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOETHERICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 21..2709
; US-08-480-145-1

Query Match 0.5%; Score 17; DB 1; Length 4165;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2811 atttagtgcatcttt 2827
|||||
Db 4070 ATTTAGTGCAATTTT 4054

RESULT 39

US-08-477-389-1/C
; Sequence 1, Application US/08477389
; Patent No. 5872219
; GENERAL INFORMATION:
; APPLICANT: Diflore, Pier P
; APPLICANT: Fazioli, Francesca
; TITLE OF INVENTION: A Substrate for the Epidermal Growth
; TITLE OF INVENTION: Factor Receptor Kinase
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: United States of America
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477.389
FILING DATE: 07-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/095.737
FILING DATE: 22-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned A
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH060.001A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4165 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 21..2709
US-08-477-389-1

Query Match 0.5%; Score 17; DB 2; Length 4165;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2811 atttagtgatctttt 2827
|||||

Db 4070 attttagtgcatTTTT 4054

RESULT 40
US-08-559-303B-72
Sequence 72, Application US/08559303B
Patent No. 5624501
GENERAL INFORMATION:
APPLICANT: NATHAN A. ELLIS, JAMES GERMAN, AND JOANNA
APPLICANT: GRODEN
TITLE OF INVENTION: METHODS FOR DIAGNOSIS AND TREATMENT
NUMBER OF INVENTION: OF BLOOM'S SYNDROME
NUMBER OF SEQUENCES: 78
CORRESPONDENCE ADDRESS:
ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN
STREET: 90 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: U.S.A.
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE DISKETTE
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/559.303B
FILING DATE: NOVEMBER 15, 1995
ATTORNEY/AGENT INFORMATION:
NAME: ELIZABETH A. BOGOSIAN
REGISTRATION NUMBER: 39,911
REFERENCE/DOCKET NUMBER: 63475/65
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-5995
TELEFAX: (212) 286-0854 or 286-0082

TELEX: TWX 710-581-4766
INFORMATION FOR SEQ ID NO: 72:
SEQUENCE CHARACTERISTICS:
LENGTH: 4437
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE:
DESCRIPTION: OTHER NUCLEIC ACID
HYPOTHETICAL: YES
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
US-08-559-303B-72

Query Match 0.5%; Score 17; DB 1; Length 4437;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3052 aatttgatgatgatgat 3068
|||||

Db 943 aatttgatgatgatgat 959

RESULT 41
US-09-175-828-72
Sequence 72, Application US/09175828
Patent No. 6221643
GENERAL INFORMATION:
APPLICANT: NATHAN A. ELLIS, JAMES GERMAN, AND JOANNA
APPLICANT: GRODEN
TITLE OF INVENTION: METHODS FOR DIAGNOSIS AND TREATMENT
NUMBER OF INVENTION: OF BLOOM'S SYNDROME
NUMBER OF SEQUENCES: 78
CORRESPONDENCE ADDRESS:
ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN
STREET: 90 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: U.S.A.
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE DISKETTE
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/175.828
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/559.303
FILING DATE: NOVEMBER 15, 1995
ATTORNEY/AGENT INFORMATION:
NAME: ELIZABETH A. BOGOSIAN
REGISTRATION NUMBER: 39,911
REFERENCE/DOCKET NUMBER: 63475/65
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-5995
TELEFAX: (212) 286-0854 or 286-0082
INFORMATION FOR SEQ ID NO: 72:
SEQUENCE CHARACTERISTICS:
LENGTH: 4437
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE:
DESCRIPTION: OTHER NUCLEIC ACID
HYPOTHETICAL: YES

ANTI-SENSE: NO
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
US-09-175-828-72

Query Match 0.5%; Score 17; DB 4; Length 4437;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3052 aattgatgatgatgat 3068
|||||
DB 943 AATTGATGATGATGAT 959

RESULT 42
US-09-103-840A-1/C
Sequence 1, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 1
LENGTH: 441529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37RV
US-09-103-840A-1

Query Match 0.5%; Score 17; DB 4; Length 441529;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2400 aacctgaagaagatgct 2416
|||||
DB 3725573 AACCTGAAGAGATGTG 3725557

RESULT 43
US-08-137-117D-48/C
Sequence 48, Application US/08137117D
Patent No. 5795965
GENERAL INFORMATION:
APPLICANT: TSUCHIYA, Masayuki
APPLICANT: SATO, Koh
APPLICANT: BENDIG, Mary
APPLICANT: JONES, Steven
APPLICANT: SALDANHA, Jose
TITLE OF INVENTION: RESHAPED HUMAN ANTIBODY TO HUMAN
NUMBER OF SEQUENCES: 158
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/137,117D
FILING DATE: 20-DEC-1993
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/JP92/00544
FILING DATE: 24-APR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 4-32084
FILING DATE: 19-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 3-95476
FILING DATE: 25-APR-1991
ATTORNEY/AGENT INFORMATION:
NAME: WEGNER, Harold C.
REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 53466/126/AOK
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 36 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-137-117D-48

Query Match 0.5%; Score 16; DB 1; Length 36;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2182 ctctcgtggtggaagc 2197
|||||
DB 22 CTCTCATGTGTGAAGC 7

RESULT 44
US-08-436-717-48/C
Sequence 48, Application US/08436717
Patent No. 5817790
GENERAL INFORMATION:
APPLICANT: TSUCHIYA, Masayuki
APPLICANT: SATO, Koh
APPLICANT: BENDIG, Mary
APPLICANT: JONES, Steven
APPLICANT: SALDANHA, Jose
TITLE OF INVENTION: RESHAPED HUMAN ANTIBODY TO HUMAN
NUMBER OF SEQUENCES: 158
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/436,717
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/137,117

FILING DATE: 20-DEC-1993
APPLICATION NUMBER: WO PCT/JP92/00544
FILING DATE: 24-APR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 4-32084
FILING DATE: 19-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 3-95476
FILING DATE: 25-APR-1991
ATTORNEY/AGENT INFORMATION:
NAME: WEGNER, Harold C.
REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 53466/126/AAOK
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 36 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-436-717-48

Query Match 0.5%; Score 16; DB 1; Length 36;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2182 ctctcatgtgtgaagc 2197
|||||

DB 22 CTCCTCATGTGTGAAGC 7

RESULT 45
US-08-943-731-153/c
Sequence 153, Application US/08943731
Patent No. 6265157
GENERAL INFORMATION:
APPLICANT: PROCKOP, DARWIN J.
APPLICANT: SPOTILA, LORETTA D.
APPLICANT: DELTAS, CONSTATINOS D.
APPLICANT: SEREDA, LARISSA
APPLICANT: LARSON, ANDREA W.
APPLICANT: PACE, MICHAEL
APPLICANT: COLIGE, ALAIN
APPLICANT: EARLY, JAMES
APPLICANT: KORRKO, JARMO
APPLICANT: ALA-KORRKO, LEENA, et al.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING
TITLE OF INVENTION: ALTERED TYPE I OR TYPE IX COLLAGEN GENE SEQUENCES
NUMBER OF SEQUENCES: 666
CORRESPONDENCE ADDRESS:
ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
STREET: ONE COMMERCE SQUARE, 2005 MARKET STREET, 22ND
STREET: FLR.
CITY: PHILADELPHIA
STATE: PA
COUNTRY: USA
ZIP: 19103-7086
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/943,731
FILING DATE: 03-OCT-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/212,322
FILING DATE: 14-MAR-1994

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/803,628
FILING DATE: 03-DEC-1991
ATTORNEY/AGENT INFORMATION:
NAME: DOYLE LEARY Ph.D., KATHRYN
REGISTRATION NUMBER: 36,317
REFERENCE/DOCKET NUMBER: 9598-27
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-965-1284
TELEFAX: 215-567-2991
TELEX: 831-494
INFORMATION FOR SEQ ID NO: 153:
SEQUENCE CHARACTERISTICS:
LENGTH: 84 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-943-731-153

Query Match 0.5%; Score 16; DB 4; Length 84;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1419 catgagccagagag 1434
|||||

DB 61 CATGAGCCAGAGAGAG 46

RESULT 46
US-08-479-577-1/c
Sequence 1, Application US/08479577
Patent No. 5633449
GENERAL INFORMATION:
APPLICANT: MILTON ZAITLIN, and Peter Palukaitis
TITLE OF INVENTION: Induction Of Resistance To
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: yahwak & Associates
STREET: 25 Skytop Drive
CITY: Trumbull
STATE: Connecticut
COUNTRY: USA
ZIP: 06611
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: Macintosh
OPERATING SYSTEM: MS-DOS
SOFTWARE: Microsoft Word 4.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/479,577
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/197,096
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: George M. Yahwak
REGISTRATION NUMBER: 26,824
REFERENCE/DOCKET NUMBER: CRF D-1123NM
TELECOMMUNICATION INFORMATION:
TELEPHONE: (203)268-1951
TELEFAX: (203)268-1951
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 132 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: RNA
US-08-479-577-1

Query Match 0.5%; Score 16; DB 1; Length 132;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2541 gcagtgaaatcctag 2556
|||||
DB 100 GCAGTGAATCCTAG 85

RESULT 47
US-08-756-977-1/c
; Sequence 1, Application US/08756977
; Patent No. 5945581

GENERAL INFORMATION:
APPLICANT: Zaitlin, Milton
APPLICANT: Palukaitis, Peter
TITLE OF INVENTION: INDUCTION OF RESISTANCE TO VIRAL
TITLE OF INVENTION: DISEASES IN PLANTS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
STREET: Clinton Square, P.O. Box 1051
CITY: Rochester
STATE: New York
COUNTRY: U.S.A.
ZIP: 14603

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/756,977
FILING DATE: 26-NOV-1996
CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: Goldman, Michael L.
REGISTRATION NUMBER: 30,727
REFERENCE/DOCKET NUMBER: 19603/10186

TELECOMMUNICATION INFORMATION:

TELEPHONE: (716) 263-1304

TELEFAX: (716) 263-1600

INFORMATION FOR SEQ. ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 132 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: RNA (genomic)

US-08-756-977-1

Query Match 0.5%; Score 16; DB 2; Length 132;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2541 gcagtgaaatcctag 2556
|||||
DB 100 GCAGTGAATCCTAG 85

RESULT 48
PCT-US93-05331-5/c

; Sequence 5, Application PC/TUS9305331
; GENERAL INFORMATION:

APPLICANT: Cornell Research Foundation, Inc.

TITLE OF INVENTION: Induction Of Resistance To

TITLE OF INVENTION: Virus Diseases By Transformation Of Plants With A Replicase Pc

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: Yahwak & Associates.

STREET: 25 Skytop Drive
CITY: Trumbull
STATE: Connecticut
COUNTRY: USA
ZIP: 06611

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: Macintosh

OPERATING SYSTEM: MS-DOS

SOFTWARE: Microsoft Word 4.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US93/05331

FILING DATE: 19930603

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: George W. Yahwak

REGISTRATION NUMBER: 26,824

REFERENCE/DOCKET NUMBER: CRF D-1123

TELECOMMUNICATION INFORMATION:

TELEPHONE: (203)268-1951

TELEFAX: (203)268-1951

INFORMATION FOR SEQ. ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 132 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA

PCT-US93-05331-5

Query Match 0.5%; Score 16; DB 5; Length 132;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2541 gcagtgaaatcctag 2556
|||||
DB 100 GCAGTGAATCCTAG 85

RESULT 49
US-08-621-018B-20

; Sequence 20, Application US/08621018B
; Patent No. 6060239

GENERAL INFORMATION:

APPLICANT: Stuart, Susan G.

APPLICANT: Hawkins, Phillip R.

APPLICANT: Sellhammer, Jeffrey J.

APPLICANT: Murry, Lynn E.

TITLE OF INVENTION: CELLUBREVIN HOMOLOGS

NUMBER OF SEQUENCES: 51

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: U.S.

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq Version 1.5

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/621,018B

FILING DATE: March 22, 1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/409,373

FILING DATE: March 23, 1995

ATTORNEY/AGENT INFORMATION:

NAME: CERRONE, MICHAEL C.

REGISTRATION NUMBER: 39,132

REFERENCE/DOCKET NUMBER: PF-0029-1 CIP

TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 284 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
LIBRARY: SYNORAT01
CLONE: 370165
US-08-621-018B-20

Query Match 0.5%; Score 16; DB 3; Length 284;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
2186 catggtggaagcagct 2201
|||||
Db 43 CATGCTGAGACCACT 58

RESULT 50
US-08-171-385-26
Sequence 26, Application US/08171385
Patent No. 5527884
GENERAL INFORMATION:
APPLICANT: Mary E. Russell
APPLICANT: Ulrike Utans
TITLE OF INVENTION: Mediators of Chronic Allograft
REJECTION
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: Wordperfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/171.385
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 05433/006001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 300
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-171-385-26

Query Match 0.5%; Score 16; DB 1; Length 300;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1995 aactggaagcagaat 2010
|||||
Db 282 AACTGAGACAGCAAT 297

RESULT 51
US-08-361-441B-26
Sequence 26, Application US/08361441B
Patent No. 6077948
GENERAL INFORMATION:
APPLICANT: Russell, Mary E.
APPLICANT: Utans, Ulrike
TITLE OF INVENTION: MEDIATORS OF CHRONIC ALLOGRAFT REJECTION
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/361.441B
FILING DATE: 21-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/171.385
FILING DATE: 21-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 05433/014001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 300 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-361-441B-26

Query Match 0.5%; Score 16; DB 3; Length 300;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1995 aactggaagcagaat 2010
|||||
Db 282 AACTGAGACAGCAAT 297

RESULT 52
US-08-916-576B-40
Sequence 40, Application US/08916576B
Patent No. 6171816
GENERAL INFORMATION:
APPLICANT: YU, GUO-LIANG
APPLICANT: DILON, PATRICK J.
APPLICANT: EBNER, REINHARD
APPLICANT: ENDRESS, GREGORY A.
TITLE OF INVENTION: NOVEL HUMAN GROWTH FACTORS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX, P.L.L.C.

STREET: 1100 NEW YORK AVENUE, SUITE 600
 CITY: WASHINGTON
 STATE: DC
 COUNTRY: US
 ZIP: 20005-3934
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/916,576B
 FILING DATE:
 CLASSIFICATION: 536
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 60/024,347
 FILING DATE: 23-AUG-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: STEFFE, ERIC K.
 REGISTRATION NUMBER: 36,688
 REFERENCE/DOCKET NUMBER: 1488.0500001
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 371-2600
 TELEFAX: (202) 371-2540
 INFORMATION FOR SEQ ID NO: 40:
 LENGTH: 311 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: CDNA
 US-08-916-576B-40

Query Match 0.5%; Score 16; DB 4; Length 311;
 Best Local Similarity 100.0%; Pred. No. 4.2e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1820 attgactctttaa 1835
 Db 108 ATTGACTCTTGA 123

RESULT 53
 US-08-991-789A-78
 Sequence 78, Application US/08991789A
 Patent No. 6225054
 GENERAL INFORMATION:
 APPLICANT: Fridakis, Tony N.
 Smith, John M.
 TITLE OF INVENTION: REED, STEVEN G.
 TREATMENT AND DIAGNOSIS OF BREAST CANCER
 NUMBER OF SEQUENCES: 292
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Seed IP Law Group
 STREET: 701 Fifth Avenue, Suite 6300
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/991,789A
 FILING DATE: 11-Dec-1997
 CLASSIFICATION: <Unknown>
 ATTORNEY/AGENT INFORMATION:
 NAME: Potler, Jane E. R.
 REGISTRATION NUMBER: 33,332

REFERENCE/DOCKET NUMBER: 210121.419C3
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031
 INFORMATION FOR SEQ ID NO: 78:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 355 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 SEQUENCE DESCRIPTION: SEQ ID NO: 78:
 US-08-991-789A-78

Query Match 0.5%; Score 16; DB 4; Length 355;
 Best Local Similarity 100.0%; Pred. No. 4.2e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2557 ctcaaatcttcaaa 2572
 Db 149 CTCGAATCTTCACAA 164

RESULT 54
 US-08-991-789A-248
 Sequence 248, Application US/08991789A
 Patent No. 6225054
 GENERAL INFORMATION:
 APPLICANT: Fridakis, Tony N.
 Smith, John M.
 TITLE OF INVENTION: REED, STEVEN G.
 TREATMENT AND DIAGNOSIS OF BREAST CANCER
 NUMBER OF SEQUENCES: 292
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Seed IP Law Group
 STREET: 701 Fifth Avenue, Suite 6300
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/991,789A
 FILING DATE: 11-Dec-1997
 CLASSIFICATION: <Unknown>
 ATTORNEY/AGENT INFORMATION:
 NAME: Potler, Jane E. R.
 REGISTRATION NUMBER: 33,332
 REFERENCE/DOCKET NUMBER: 210121.419C3
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031
 INFORMATION FOR SEQ ID NO: 248:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 355 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 SEQUENCE DESCRIPTION: SEQ ID NO: 248:
 US-08-991-789A-248

Query Match 0.5%; Score 16; DB 4; Length 355;
 Best Local Similarity 100.0%; Pred. No. 4.2e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 2557 ctcaaatcttcaaa 2572
 ||||||||||||||||

Db 149 CTCGATCTTCACAA 164

RESULT 55

US-08-470-179-100
Sequence 100, Application US/08470179
Patent No. 5645994

GENERAL INFORMATION:

APPLICANT: Huang Ph.D, Mai Mun

TITLE OF INVENTION: Method and Compositions for

TITLE OF INVENTION: Identification of Species in a Sample

NUMBER OF SEQUENCES: 207

CORRESPONDENCE ADDRESS:

ADDRESSEE: Trask, Britt and Rossa

STREET: P.O. Box 2550

CITY: Salt Lake City

STATE: Utah

COUNTRY: USA

ZIP: 84110

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA: US/08/470.179

APPLICATION NUMBER:

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Sweigert Ph.D, Susan E.

REGISTRATION NUMBER: 36,289

REFERENCE/DOCKET NUMBER: 2601

TELECOMMUNICATION INFORMATION:

TELEPHONE: 801-532-1922

TELEFAX: 801-531-9168

INFORMATION FOR SEQ ID NO: 100:

SEQUENCE CHARACTERISTICS:

LENGTH: 424 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: not relevant

MOLECULE TYPE: other nucleic acid

DESCRIPTION: /desc = "gyra gene segment"

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Treponema pallidum

US-08-470-179-100

Query Match

Best Local Similarity 0.5%; Score 16; DB 1; Length 424;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 14 tgggaagcttcaccc 29

Db 107 TGGGAGCTTCATCC 122

RESULT 56

US-08-943-731-127

Sequence 127, Application US/08943731

Patent No. 6265157

GENERAL INFORMATION:

APPLICANT: PROCKOP, DARWIN J.

APPLICANT: SPOTILA, LORETTA D.

APPLICANT: DELTAS, CONSTANTINOS D.

APPLICANT: SEREDA, LARISSA

APPLICANT: LARSON, ANDREA W.

APPLICANT: PACK, MICHAEL

APPLICANT: COLIGE, ALAIN

APPLICANT: EARLY, JAMES

APPLICANT: KORRKO, JARMO

APPLICANT: ALA-KORRKO, LEENA, et al.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING

TITLE OF INVENTION: ALTERED TYPE I OR TYPE IX COLLAGEN GENE SEQUENCES

NUMBER OF SEQUENCES: 666

CORRESPONDENCE ADDRESS:

ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.

STREET: ONE COMMERCE SQUARE, 2005 MARKET STREET, 22ND

STATE: FLR.

CITY: PHILADELPHIA

STATE: PA

COUNTRY: USA

ZIP: 19103-7086

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA: US/08/943.731

APPLICATION NUMBER: US/08/943.731

FILING DATE: 03-OCT-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/212,322

FILING DATE: 14-MAR-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/803,628

FILING DATE: 03-DEC-1991

ATTORNEY/AGENT INFORMATION:

NAME: DOYLE LEARY Ph.D., KATHRYN

REGISTRATION NUMBER: 36,317

REFERENCE/DOCKET NUMBER: 9598-27

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-965-1284

TELEFAX: 215-567-2991

INFORMATION FOR SEQ ID NO: 127:

SEQUENCE CHARACTERISTICS:

LENGTH: 434 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-943-731-127

Query Match

Best Local Similarity 0.5%; Score 16; DB 4; Length 434;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1979 tgatccttgctctc 1994

Db 196 TGTATCTTGTCTCTC 211

RESULT 57

US-08-899-786-14/c

Sequence 14, Application US/08899786

Patent No. 6001572

GENERAL INFORMATION:

APPLICANT: Toothman, Penelope

TITLE OF INVENTION: Method of Identifying Aloe Using

NUMBER OF SEQUENCES: 32

CORRESPONDENCE ADDRESS:

ADDRESSEE: Swanson & Bratschun, L.L.C.

STREET: 8400 E. Prentice Avenue, Suite 200

CITY: Englewood

STATE: Colorado

COUNTRY: USA

ZIP: 80111

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3 1/2 diskette, 1.44 MB

OPERATING SYSTEM: MS-DOS

SOFTWARE: Wordperfect 8.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/899,786
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/022,611
FILING DATE: 26-JULY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Barry J. Swanson
REGISTRATION NUMBER: 33,215
REFERENCE/DOCKET NUMBER: UNI.07
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 793-3333
TELEFAX: (303) 793-3433
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 520 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE:
US-08-899-786-14

Query Match 0.5%; Score 16; DB 3; Length 520;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 333 gattggctcaggatt 348
|||||
DB 98 gattggctcaggatt 83

RESULT 58
US-08-899-786-16/c
Sequence 16, Application US/08899786
Patent No. 6001572
GENERAL INFORMATION:
APPLICANT: Toothman, Penelope
TITLE OF INVENTION: Method of Identifying Aloe Using
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Swanson & Bratschun, L.L.C.
STREET: 8400 E. Prentice Avenue, Suite 200
CITY: Englewood
STATE: Colorado
COUNTRY: USA
ZIP: 80111
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3 1/2 diskette, 1.44 MB
COMPUTER: IBM PC compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 8.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/899,786
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/022,611
FILING DATE: 26-JULY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Barry J. Swanson
REGISTRATION NUMBER: 33,215
REFERENCE/DOCKET NUMBER: UNI.07
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 793-3333
TELEFAX: (303) 793-3433
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 362 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE:
US-08-899-786-16

Query Match 0.5%; Score 16; DB 3; Length 562;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 333 gattggctcaggatt 348
|||||
DB 100 gattggctcaggatt 85

RESULT 59
US-08-906-156A-67/c
Sequence 67, Application US/08906156A
Patent No. 6287854
GENERAL INFORMATION:
APPLICANT: SPURR, NIGEL K
APPLICANT: GRAY, IAN C
APPLICANT: STEWART, LORNA M
TITLE OF INVENTION: DIAGNOSIS OF SUSCEPTIBILITY TO CANCER
TITLE OF INVENTION: AND TREATMENT THEREOF
NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHAYE P. C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22201
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/906,156A
FILING DATE: 05-AUG-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/042,655
FILING DATE: 02-APR-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/033,147
FILING DATE: 13-DEC-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/005,840
FILING DATE: 23-OCT-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/96GB/02588
FILING DATE: 22-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B.J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 1090-14
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-816-4000
TELEFAX: 703-816-4100
INFORMATION FOR SEQ ID NO: 67:
SEQUENCE CHARACTERISTICS:
LENGTH: 573 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: human
US-08-906-156A-67

Query Match 0.5%: Score 16; DB 4; Length 573;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 196 tcattcacatgatattt 211
|||||
Db 76 TCATTACATGATTTT 61

RESULT 60

US-08-899-786-15/c
; Sequence 15, Application US/08899786
; Patent No. 6001572

GENERAL INFORMATION:

APPLICANT: Toothman, Penelope
TITLE OF INVENTION: Method of Identifying Aloe Using

NUMBER OF SEQUENCES: 32

CORRESPONDENCE ADDRESS:
ADDRESSEE: Swanson & Bratschun, L.L.C.
STREET: 6400 E. Prentice Avenue, Suite 200
CITY: Englewood
STATE: Colorado
COUNTRY: USA

ZIP: 80111

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3 1/2 diskette, 1.44 MB

COMPUTER: IBM pc compatible

OPERATING SYSTEM: MS-DOS

SOFTWARE: Wordperfect 8.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/899,786

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/022,611

FILING DATE: 26-JULY-1996

ATTORNEY/AGENT INFORMATION:

NAME: Barry J. Swanson

REGISTRATION NUMBER: 33,215

REFERENCE/DOCKET NUMBER: UNI.07

TELECOMMUNICATION INFORMATION:

TELEPHONE: (303) 793-3333

TELEFAX: (303) 793-3433

INFORMATION FOR SEQ ID NO: 15:

SEQUENCE CHARACTERISTICS:

LENGTH: 574 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE:

US-08-899-786-15

Query Match 0.5%: Score 16; DB 3; Length 574;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 333 gatttgctcagatt 348
|||||
Db 112 GATTGGCTCAGATT 97

RESULT 61

US-08-906-156A-19/c

; Sequence 19, Application US/08906156A
; Patent No. 6287854

GENERAL INFORMATION:

APPLICANT: SPURR, NIGEL K

APPLICANT: GRAY, IAN C

APPLICANT: STEWART, LORNA M

TITLE OF INVENTION: DIAGNOSIS OF SUSCEPTIBILITY TO CANCER

TITLE OF INVENTION: AND TREATMENT THEREOF

NUMBER OF SEQUENCES: 94

CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHAYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VA
COUNTRY: USA

ZIP: 22201

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/906,156A

FILING DATE: 05-AUG-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/042,655

FILING DATE: 02-APR-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/005,840

FILING DATE: 23-OCT-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/96GB/02588

FILING DATE: 22-OCT-1996

ATTORNEY/AGENT INFORMATION:

NAME: SADOFF, B.J.

REGISTRATION NUMBER: 36,663

REFERENCE/DOCKET NUMBER: 1090-14

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-816-4000

TELEFAX: 703-816-4100

INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:

LENGTH: 590 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

ORIGINAL SOURCE:

ORGANISM: Part of gene corresponding to IMAGE 264611

US-08-906-156A-19

Query Match 0.5%: Score 16; DB 4; Length 590;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 196 tcattcacatgatattt 211
|||||
Db 76 TCATTACATGATTTT 61

RESULT 62

US-09-385-982-233

; Sequence 233, Application US/09385982
; Patent No. 6262334

GENERAL INFORMATION:

APPLICANT: ENDEGE, WILSON O., ET AL.

TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION

TITLE OF INVENTION: PRODUCTS: II

FILE REFERENCE: CCDNA-260XX

CURRENT APPLICATION NUMBER: US/09/385,982

EARLIER FILING DATE: 1999-08-30

EARLIER APPLICATION NUMBER: 09/328,111

EARLIER FILING DATE: 1999-06-08

EARLIER APPLICATION NUMBER: 60/117,393

EARLIER FILING DATE: 1999-01-27

EARLIER APPLICATION NUMBER: 60/098,639

Tue Mar 26 11:35:52 2002

```
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 233
; LENGTH: 603
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(603)
; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-233
```

```
Query Match          0.5%; Score 16; DB 4; Length 603;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy 2412 atgcttatttcatt 2427
      |||||
Db 373 atgcttatttcatt 388
```

```
RESULT 63
US-09-328-111-205/c
; Sequence 205, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Derfl, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; CURRENT FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 205
; LENGTH: 607
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(607)
; OTHER INFORMATION: n = A,T,C or G
US-09-328-111-205
```

```
Query Match          0.5%; Score 16; DB 4; Length 607;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy 2977 cagcattatcagaaga 2992
      |||||
Db 325 CAGCATTATCAGAGAA 310
```

```
RESULT 64
US-09-385-982-101/c
; Sequence 101, Application US/09385982
; Patent No. 6262334
```

```
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCDNA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; CURRENT FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 101
; LENGTH: 625
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(625)
; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-101
```

```
Query Match          0.5%; Score 16; DB 4; Length 625;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy 763 ttcttctctcttga 778
      |||||
Db 159 ttcttctctcttga 144
```

```
RESULT 65
US-08-468-347-25/c
; Sequence 25, Application US/08468347
; Patent No. 5783421
; GENERAL INFORMATION:
; APPLICANT: Zeelon, Elisha P.
; APPLICANT: Werber, Moshe M.
; APPLICANT: Levanon, Avigdor
; TITLE OF INVENTION: NOVEL POLYPEPTIDE HAVING FACTOR Xa
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,347
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/225,442
; FILING DATE: 08-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 0317/43020-A/JPW/EAB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-664-0525
; TELEFAX: 212-664-0525
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
```

LENGTH: 697 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-468-347-25

Query Match 0.5%; Score 16; DB 1; Length 697;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2940 tttttgacttagta 2955
|||||
Db 678 TTTTGTGACTTACTGA 663

RESULT 66
US-08-467-389-25/C
Sequence 25, Application US/08467389
Patent No. 5824641
GENERAL INFORMATION:
APPLICANT: Zeelon, Elisha P.
APPLICANT: Werber, Moshe M.
APPLICANT: Levanon, Avigdor
TITLE OF INVENTION: NOVEL POLYPEPTIDE HAVING FACTOR Xa
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/467,389
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/225,442
FILING DATE: 08-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 0317/43020-A/JPW/EAB
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-977-9550
TELEFAX: 212-664-0525
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 697 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-467-389-25

Query Match 0.5%; Score 16; DB 1; Length 697;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2940 tttttgacttagta 2955
|||||
Db 678 TTTTGTGACTTACTGA 663

RESULT 67
US-08-779-379-25/C
Sequence 25, Application US/08779379
Patent No. 5858970
GENERAL INFORMATION:
APPLICANT: Zeelon, Elisha P.
APPLICANT: Werber, Moshe M.
APPLICANT: Levanon, Avigdor
TITLE OF INVENTION: NOVEL POLYPEPTIDE HAVING FACTOR Xa
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/779,379
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/225,442
FILING DATE: 08-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 0317/43020-A/JPW/EAB
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-977-9550
TELEFAX: 212-664-0525
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 697 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-779-379-25

Query Match 0.5%; Score 16; DB 2; Length 697;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2940 tttttgacttagta 2955
|||||
Db 678 TTTTGTGACTTACTGA 663

RESULT 68
US-08-469-219-25/C
Sequence 25, Application US/08469219
Patent No. 5863534
GENERAL INFORMATION:
APPLICANT: Zeelon, Elisha P.
APPLICANT: Werber, Moshe M.
APPLICANT: Levanon, Avigdor
TITLE OF INVENTION: NOVEL POLYPEPTIDE HAVING FACTOR Xa

```

1  TITLE OF INVENTION:  INHIBITORY ACTIVITY
2  NUMBER OF SEQUENCES:  25
3  CORRESPONDENCE ADDRESS:
4  ADDRESSEE:  Cooper & Dunham
5  STREET:  30 Rockefeller Plaza
6  CITY:  New York
7  STATE:  New York
8  ZIP:  10112
9  COMPUTER READABLE FORM:
10 MEDIUM TYPE:  Floppy disk
11 COMPUTER:  IBM PC compatible
12 OPERATING SYSTEM:  PC-DOS/MS-DOS
13 SOFTWARE:  PatentIn Release #1.0, Version #1.25
14 CURRENT APPLICATION DATA:
15 APPLICATION NUMBER:  US/08/469,219
16 FILING DATE:
17 CLASSIFICATION:  435
18 PRIOR APPLICATION DATA:
19 APPLICATION NUMBER:  08/225,442
20 FILING DATE:  08-Apr-1994
21 ATTORNEY/AGENT INFORMATION:
22 NAME:  White, John P.
23 REGISTRATION NUMBER:  28,678
24 REFERENCE/DOCKET NUMBER:  0317/43020-A/JPM/EAB
25 TELECOMMUNICATION INFORMATION:
26 TELEPHONE:  212-977-9550
27 TELEFAX:  212-664-0525
28 INFORMATION FOR SRO ID NO:  25:
29 SEQUENCE CHARACTERISTICS:
30 LENGTH:  697 base pairs
31 TYPE:  nucleic acid
32 STRANDEDNESS:  single
33 TOPOLOGY:  linear
34 MOLECULE TYPE:  DNA (genomic)
35 HYPOTHETICAL:  NO
36 ANTI-SENSE:  NO
37 FRAGMENT TYPE:  N-terminal
38
39 JS-08-469-219-25

```

```

Query Match Similarity      0.5%; Score 16; DB 2; Length 697;
Best Local Smilarity       100.0%; Pred. No. 4,2e+02;
Matches    16; Conservative    0; Mismatches    0; Indels    0; Gaps    0

OY      2940   tttttgaccttagta 2955
              ||| ||||| ||||| |||
Db       678   TTTTGTGACTTAGTA 663

SUTL     69
-09-228-152-25/c
Sequence 25, Application US/09228152
Patent No. 6211341
GENERAL INFORMATION:
APPLICANT: Zeelon, Elisha P.
APPLICANT: Werber, Moshe M.
APPLICANT: Levanon, Avigdor
TITLE OF INVENTION: NOVEL POLYPEPTIDE HAVING FACTOR Xa INHIBITORY ACTIVITY
FILE REFERENCE: 430208ya
CURRENT APPLICATION NUMBER: US/09/228,152
CURRENT FILING DATE: 1999-01-11
NUMBER OF SEQ ID NOS: 25
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 25
LENGTH: 697
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: DNA sequence
US-09-228-152-25
```

Query Match	Score 16:	DB 4:	Length 697:
Best Local Similarity	100.0%:	Pred. No. 4.2e+02:	
Matches	16:	Conservative	0: Mismatches 0: Indels 0: Gaps 0:
Oy	2940	tttttgactttagta	2955
Db	678	tttttgactttagta	663

```

RESULT 70
US-08-916-576B-5/c
: Sequence 5, Application US/08916576B
: Patent No. 6171816
: GENERAL INFORMATION:
: APPLICANT: YU, GUO-LIANG
: APPLICANT: DILLON, PATRICK J.
: APPLICANT: EBNER, REINHARD
: APPLICANT: ENDRESS, GREGORY A.
: TITLE OF INVENTION: NOVEL HUMAN GROWTH FACTORS
: NUMBER OF SEQUENCES: 45
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: STERNE, KESSLER, GOLDSSTEIN & FOX, P. L. L. C.
: STREET: 1100 NEW YORK AVENUE, SUITE 600
: CITY: WASHINGTON
: STATE: DC
: COUNTRY: US
: ZIP: 20005-3934
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/916,576B
: FILING DATE:
: CLASSIFICATION: 536
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/024,347
: FILING DATE: 23-AUG-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: STEFFE, ERIC K.
: REGISTRATION NUMBER: 36,688
: REFERENCE/DOCKET NUMBER: 1488..0500001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 371-2600
: TELEFAX: (202) 371-2540
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 732 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 49..546
: FEATURE:
: NAME/KEY: mat_peptide
: LOCATION: 118..546
: FEATURE:
: NAME/KEY: sig_peptide
: LOCATION: 49..117
: US-08-916-576B-5

Query Match 0.5%; Score 16; DB 4; Length 732;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

RESULT 71
US-08-152-485-3
; Sequence 3, Application US/08152485
; Patent No. 5539094
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; APPLICANT: Takayama, Shinichi
; APPLICANT: Sato, Takaaki
; TITLE OF INVENTION: Bcl-2-associated proteins
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CAMPBELL AND FLORES
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/152,485
; FILING DATE: 10-NOV-1993
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 9725
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 733 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-152-485-3

Query Match 0.5%; Score 16; DB 1; Length 733;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2732 gttaaacatttgag 2747
|||||

DB 229 GTTGAACATTGGAG 244

RESULT 72
US-08-463-089-3
; Sequence 3, Application US/08463089
; Patent No. 5641866
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; APPLICANT: Takayama, Shinichi
; APPLICANT: Sato, Takaaki
; TITLE OF INVENTION: Bcl-2-associated proteins
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CAMPBELL AND FLORES
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,089
; FILING DATE: 10-NOV-1993
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 9725
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 733 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-463-089-3

Query Match 0.5%; Score 16; DB 1; Length 733;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2732 gttaaacatttgag 2747
|||||

DB 229 GTTGAACATTGGAG 244

RESULT 73
US-08-461-360A-3
; Sequence 3, Application US/08461360A
; Patent No. 5650491
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; APPLICANT: Takayama, Shinichi
; APPLICANT: Sato, Takaaki
; TITLE OF INVENTION: Bcl-2-associated proteins
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CAMPBELL AND FLORES
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,360A
; FILING DATE: 10-NOV-1993
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 9725
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 733 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-461-360A-3

Query Match 0.5%; Score 16; DB 1; Length 733;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2732 gttgaacattggag 2747
|||||
DB 229 GTTGAACAATTGGAG 244

RESULT 74

US-08-461-359-3
; Sequence 3, Application US/08461359
; Patent No. 5686595
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; APPLICANT: Takayama, Shinichi
; APPLICANT: Sato, Takaaki
; TITLE OF INVENTION: Bcl-2-associated proteins
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CAMPBELL AND FLORES
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,359
; FILING DATE: 10-NOV-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 9725
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 733 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; S-08-461-359-3

Query Match 0.5%; Score 16; DB 1; Length 733;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2732 gttgaacattggag 2747
|||||
DB 229 GTTGAACAATTGGAG 244

RESULT 75

PCT-US94-12904-3
; Sequence 3, Application PC/TUS9412904
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; APPLICANT: Takayama, Shinichi
; APPLICANT: Sato, Takaaki
; TITLE OF INVENTION: Bcl-2-associated proteins
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:

ADDRESSEE: CAMPBELL AND FLORES
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/12904
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/152,485
; FILING DATE: 12-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1201
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 733 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; PCT-US94-12904-3

Query Match 0.5%; Score 16; DB 5; Length 733;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2732 gttgaacattggag 2747
|||||
DB 229 GTTGAACAATTGGAG 244

Search completed: March 25, 2002, 18:01:41
Job time: 16640 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW nucleic - nucleic search, using sw model

Run on: March 25, 2002, 14:18:56 : Search time 270.41 Seconds

(without alignments)
9933.064 Million cell updates/sec

Title: US-09-697-089-1

Perfect score: 3133
Sequence: 1 cgcctcagccgcggtgggaag.....aagtgctactgaagccagta 3133

Scoring table: OLIGO_NUC
Gapop 60.0, Gapext 60.0

Searched: 930621 seqs, 428662619 residues

Word size: 0

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database: N.Geneseq_1101.*

1: /SID2/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SID2/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SID2/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SID2/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SID2/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SID2/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SID2/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SID2/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SID2/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SID2/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SID2/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SID2/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SID2/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SID2/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SID2/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SID2/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SID2/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SID2/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SID2/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SID2/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SID2/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SID2/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3133	100.0	3133	22	Human caspase recr
2	2969	94.8	3545	22	Mutine EST-derived
3	2634	84.1	3615	22	Human caspase recr
4	1807	57.7	2950	22	Human protein enco
5	515	16.4	2735	22	Human colon cancer
6	242	7.7	421	22	Probe #4322 for ge
7	242	7.7	421	22	Probe #4450 used t
8	242	7.7	421	22	Probe #4204 used t
9	220	7.0	220	22	Probe #13523 for g
10	220	7.0	220	22	Probe #17590 used
11	220	7.0	22	22	Probe #9197 used t

12	19	0.6	549	22	AAH11452	Human CDNA clone (
13	19	0.6	579	19	AA330628	H. pylori cell env
14	19	0.6	649	19	AA114445	H. pylori GHP0 875
15	19	0.6	1908	21	AA13098	Aspergillus oryzae
16	19	0.6	2825	22	AAH16202	Human CDNA sequenc
17	18	0.6	116	21	AAC14986	Human secreted pro
18	18	0.6	274	21	AAA31261	Plant microsateili
19	18	0.6	322	16	AA224250	Human gene signatu
20	18	0.6	355	21	AAA31416	Plant microsateili
21	18	0.6	377	21	AAA31325	Plant microsateili
22	18	0.6	435	21	AAAC09333	Human secreted pro
23	18	0.6	446	21	AAAC09333	Human secreted pro
24	18	0.6	454	22	AA114905	Probe #4838 for ge
25	18	0.6	454	22	AA136257	Probe #4913 used t
26	18	0.6	454	22	AA104685	Probe #4676 used t
27	18	0.6	466	21	AAAC01809	Human secreted pro
28	18	0.6	468	21	AA331287	Plant microsateili
29	18	0.6	472	22	AA111787	Probe #1720 for ge
30	18	0.6	472	22	AA133101	Probe #1787 used t
31	18	0.6	472	22	AA101718	Probe #1709 used t
32	18	0.6	594	11	AAO05868	Sequence encoding
33	18	0.6	652	21	AAE07589	Fusarium venenatum
34	18	0.6	682	21	AAAC44321	Arabidopsis thalia
35	18	0.6	814	22	AAH05406	Human CDNA clone (
36	18	0.6	888	20	AA220212	Enterococcus faeca
37	18	0.6	960	22	AAH67086	C glutamicum codin
38	18	0.6	992	11	AAO05870	Sequence encoding
39	18	0.6	1034	21	AAE25257	Human secreted pro
40	18	0.6	1083	22	AAE71387	Corynbacterium g1
41	18	0.6	1205	20	AA242236	Human normal blood
42	18	0.6	1250	9	AAH03316	Transcription cont
43	18	0.6	1473	21	AAAC40012	Arabidopsis thalia
44	18	0.6	1491	21	AAH47150	DNA encoding a ser
45	18	0.6	1515	19	AAV40737	C. fells esterase,
46	18	0.6	1515	19	AAV40738	C. fells esterase,
47	18	0.6	1561	21	AAE21032	Human low adenosin
48	18	0.6	1561	21	AAA34910	Human adenosine re
49	18	0.6	1579	22	AAH13702	Human CDNA sequenc
50	18	0.6	1611	12	AAO12528	Thymidylate phosph
51	18	0.6	1611	21	AAH47151	DNA encoding a ser
52	18	0.6	1619	21	AAH15908	Human protein clon
53	18	0.6	1650	19	AAV40760	C. fells esterase,
54	18	0.6	1650	19	AAV40761	C. fells esterase,
55	18	0.6	1723	20	AAV41287	Human vesicular bi
56	18	0.6	1800	15	AAO67601	Retinoblastoma pro
57	18	0.6	1878	22	AAE25374	Nucleotide sequenc
58	18	0.6	1926	20	AAV82522	Sphingomonas capsu
59	18	0.6	1962	20	AAZ77504	Human ovarian tumo
60	18	0.6	1982	19	AAV40735	C. fells esterase,
61	18	0.6	1982	19	AAV40736	C. fells esterase,
62	18	0.6	2144	19	AAV40758	C. fells esterase,
63	18	0.6	2144	19	AAV40759	C. fells esterase,
64	18	0.6	2299	19	AAZ96326	S. pneumoniae derl
65	18	0.6	2299	19	AAV42995	S. pneumoniae pneu
66	18	0.6	3000	20	AAH84399	S. capsulata IPO12
67	18	0.6	3001	21	AAH51776	Chromosome 13q31-q
68	18	0.6	3001	21	AAH51787	Chromosome 13q31-q
69	18	0.6	3275	22	AAE25373	Genomic sequence o
70	18	0.6	3529	21	AAH31144	Human colon cancer
71	18	0.6	3942	20	AAV80604	Kidney injury asso
72	18	0.6	3969	21	AAH44473	DNA encoding an as
73	18	0.6	4215	22	AA158415	Human polynucleoti
74	18	0.6	4671	22	AA160201	Human polynucleoti
75	18	0.6	4851	20	AAH13198	Human polyunsatur
76	18	0.6	4982	18	AAH85267	Enterococcus faeca
77	18	0.6	5212	9	AAH80317	HMC-CoA reductase
78	18	0.6	5292	21	AAE21374	Transcription cont
79	18	0.6	5292	21	AAH35552	Human low adenosin
80	18	0.6	5511	17	AAH41853	Human adenosine re
81	18	0.6	6212	21	AAE21375	CDNA encoding Plas
82	18	0.6	6312	21	AAH35253	Human low adenosin
83	18	0.6	7380	20	AAH4028	Human adenosine re
84	18	0.6	7383	22	AAH00153	MP9 promoter beta
						Matrix metalloprot

```
85 18 0.6 7386 22 AAI57671 Human colorectal c
86 18 0.6 45186 22 AAF60478 Wld-type human CT
87 18 0.6 236303 22 AAS11614 Human genomic DNA
88 18 0.6 349980 22 AAH68530 C glutamincum codin
89 18 0.6 1664976 19 AAV21209 Methanococcus jann
90 17 0.5 21 20 AA218259 p450 enzyme gene s
91 17 0.5 47 21 AA267938 Human map-related
92 17 0.5 51 21 AA299916 Sequence of the st
93 17 0.5 121 21 AAC10211 Human secreted pro
94 17 0.5 212 21 AAC13747 Human secreted pro
95 17 0.5 276 20 AAH8937 EST clone HM846.
96 17 0.5 290 20 AA222458 Internal Transcrib
97 17 0.5 312 22 AAH66827 C glutamincum codin
98 17 0.5 333 21 AAC22687 Human secreted exp
99 17 0.5 359 21 AAA44118 Human secreted exp
c 100 17 0.5 383 18 AAH83360 Breast cancer tumo
```

ALIGNMENTS

```
SULT 1
ID AS03945 standard; CDNA: 3133 BP.
```

```
12-SEP-2001 (first entry)
```

```
Human caspase recruitment domain 12 (CARD-12) CDNA.
```

```
KW Caspase recruitment domain; CARD-12; apoptosis; stress-related pathway;
KW cancer; viral infection; poxvirus; adenovirus; autoimmune disorder;
KW systemic lupus erythematosus; arthritis; neurological disorder; stroke;
KW Alzheimer's disease; amyotrophic lateral sclerosis; haematologic disease;
KW aplastic anaemia; myocardial infarction; inflammatory disorder;
KW Crohn's disease; insulin-dependent diabetes; contact dermatitis;
KW psoriasis; graft rejection; bacterial infection; lepromatous leprosy;
KW tuberculosis; ischaemic brain injury; hypoxic brain injury; ss;
KW kidney ischaemia; reperfusion injury; acute bacterial meningitis;
KW excitotoxic brain damage; liver disease.
```

```
XX Homo sapiens.
```

```
XX OS Location/Qualifiers
XX FH Key 36..3110
XX FT CDS /*tag= a
XX FT /product= "Human CARD-12"
```

```
XX MO200130971-A2.
```

```
XX 03-MAY-2001.
```

```
XX 26-OCT-2000; 2000WO-US29643.
```

```
XX 27-OCT-1999; 99US-0161822.
```

```
XX (MILL-) MILLENNIUM PHARM INC.
```

```
XX PA Bertin J, Robison KE;
```

```
XX PI WPI; 2001-308628/32.
```

```
XX DR P-PSDB; AAU02880.
```

```
XX Isolated caspase recruitment domain-12 polypeptide and nucleic acids
```

```
XX encoding them, useful for treating and diagnosing disorders associated
```

```
XX with abnormal apoptosis such as cancer, arthritis and Alzheimer's
```

```
XX disease -
```

```
XX Claim 2; Fig 1; 93pp; English.
```

```
XX The sequence represents a CDNA which encodes the human caspase
```

```
XX recruitment domain 12 (CARD-12) polypeptide. CARD domains are found in a
```

```
CC number of proteins that transmit signals that activate apoptosis and
CC inflammatory pathways in response to stress and other stimuli. Therefore,
CC CARD-12 and its corresponding nucleic acid may be used in treatment and
CC diagnosis of patients suffering from disorders associated with an
CC abnormal level (an increase or a decrease) of apoptotic cell death or
CC abnormal activity of stress-related pathways. The disorders include
CC cancer, viral infections (e.g. caused by poxviruses, adenoviruses),
CC autoimmune disorders (e.g. systemic lupus erythematosus, arthritis),
CC neurological disorders (e.g. Alzheimer's disease, amyotrophic lateral
CC sclerosis), haematologic diseases (e.g. aplastic anaemia, myocardial
CC infarction, stroke), inflammatory and immune system disorders (e.g.
CC Crohn's disease, insulin-dependent diabetes, contact dermatitis,
CC psoriasis, graft rejection), bacterial infections (e.g. tuberculosis,
CC lepromatous leprosy), ischaemic and hypoxic brain injury, kidney
CC ischaemia/reperfusion injury, excitotoxic brain damage, acute bacterial
CC meningitis and liver disease.
```

```
CC Sequence 3133 BP; 903 A; 691 C; 729 G; 810 T; 0 other;
```

```
Query Match 100.0%; Score 3133; DB 22; Length 3133;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3133; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 cgcctagccggttggaagcttccacagaacaatgcaattcataaggacatagcc 60
DB 1 cgcctagccggttggaagcttccacagaacaatgcaattcataaggacatagcc 60
OY 61 gagccctatcaagaatggaatgactgtttataagaacaatcacatgacctattg 120
DB 61 gagccctatcaagaatggaatgactgtttataagaacaatcacatgacctattg 120
OY 121 tatgaatgttcgtgatcgcaagaagtaaacatatttgcgcgaagaagtgagcag 180
DB 121 tatgaatgttcgtgatcgcaagaagtaaacatatttgcgcgaagaagtgagcag 180
OY 121 tatgaatgttcgtgatcgcaagaagtaaacatatttgcgcgaagaagtgagcag 180
DB 121 tatgaatgttcgtgatcgcaagaagtaaacatatttgcgcgaagaagtgagcag 180
OY 181 atgcgtcgaagaggaatcattcaatgatttggaaaaggttcagaagtcctgtaacct 240
DB 181 atgcgtcgaagaggaatcattcaatgatttggaaaaggttcagaagtcctgtaacct 240
OY 241 ttcttaaatcccttaagaagtggaacttcccttcttccagacttgatgacaagtc 300
DB 241 ttcttaaatcccttaagaagtggaacttcccttcttccagacttgatgacaagtc 300
OY 241 ttcttaaatcccttaagaagtggaacttcccttcttccagacttgatgacaagtc 300
DB 241 ttcttaaatcccttaagaagtggaacttcccttcttccagacttgatgacaagtc 300
OY 301 ttcttcacagacatcagaagaagagcttgacagatttgcgcagagatttaagagactgt 360
DB 301 ttcttcacagacatcagaagaagagcttgacagatttgcgcagagatttaagagactgt 360
OY 301 ttcttcacagacatcagaagaagagcttgacagatttgcgcagagatttaagagactgt 360
DB 301 ttcttcacagacatcagaagaagagcttgacagatttgcgcagagatttaagagactgt 360
OY 361 accatacccatcttcttgaaactttatcccttggtggaagatatgacattatttta 420
DB 361 accatacccatcttcttgaaactttatcccttggtggaagatatgacattatttta 420
OY 421 acttgaagaacaccttcaagaacctgtcctgttggaagaagaacacacacacacacac 480
DB 421 acttgaagaacaccttcaagaacctgtcctgttggaagaagaacacacacacacacac 480
OY 481 tggagcagctgacaccttgaaatggcctcctcagagccttcacagagccctgacatctgaag 540
DB 481 tggagcagctgacaccttgaaatggcctcctcagagccttcacagagccctgacatctgaag 540
OY 541 ggggaatctggcaagaaggaagtcacactgtcgtcgaagcattgacatgctctggggtccg 600
DB 541 ggggaatctggcaagaaggaagtcacactgtcgtcgaagcattgacatgctctggggtccg 600
OY 601 gaaagtgaagagcttgacaaagtcaattcgttcttccctcgttcacagcagggcc 660
DB 601 gaaagtgaagagcttgacaaagtcaattcgttcttccctcgttcacagcagggcc 660
OY 601 gaaagtgaagagcttgacaaagtcaattcgttcttccctcgttcacagcagggcc 660
DB 601 gaaagtgaagagcttgacaaagtcaattcgttcttccctcgttcacagcagggcc 660
OY 661 aggtggaacttttgaacacctctgtgatacaactctgataataactgacacacacaga 720
DB 661 aggtggaacttttgaacacctctgtgatacaactctgataataactgacacacacaga 720
OY 721 agcagacatcatgcatgctgctgaaagctgcgacagaaggttcttcttcttctgattg 780
DB 721 agcagacatcatgcatgctgctgaaagctgcgacagaaggttcttcttcttctgattg 780
```

[illegible]

QY	1092	gggtctccactctccacacaaacagctgtgtccatacctctctatgactctgttgatacag	1151
Db	1288	gggtctccactctccacacaaacagctgtgtccatacctctctatgactctgttgatacag	1347
QY	1152	aaaaacaaacacacataaaggctgtgtgcttgcgaagtactcttcoggaagcttgcagac	1211
Db	1348	aaaaaacaacacataaaggctgtgtgcttgcgaagtactcttcoggaagcttgcagac	1407
QY	1212	tcctggagactagctcttgagggtgtgtctccacaaagtattatttgcagtcgagagat	1271
Db	1408	tggtgatacctagctctcttgagggtgtgtctccacaaagtattatttgcagtcgagagat	1467
QY	1272	gtgtccagctgtgaatgaggatgtccctctgcacacactgtggctctctctgtataatacagct	1331
Db	1468	gtgtccagctgtgaatgaggatgtccctctgcacacactgtggctctctctgtataatacagct	1527
QY	1332	caaaagttcaagccaaagtataaatctctctcacaagtcattccaggaagtccacagcagga	1391
Db	1528	caaaagttcaagccaaagtataaatctctcacaagtcattccaggaagtccacagcagga	1587
QY	1392	cgaaagctcagcagttatttgacgtctcctcagccagaggaggctgaccaaagggaatgt	1451
Db	1588	cgaaagctcagcagttatttgacgtctcctcagccagaggaggctgaccaaagggaatgt	1647
QY	1452	tacttgcagaaaaatgtgttccatttcgcagatataccactatagcagccgtctcgg	1511
Db	1648	tacttgcagaaaaatgtgttccatttcgcagatataccactatagcagccgtctcgg	1707
QY	1512	tacacctgtgtggttcactgttcttgagaagccacagagcctgttatgtgaagcaacctcgacagctg	1571
Db	1708	tacacctgtgtggttcactgttcttgagaagccacagagcctgttatgtgaagcaacctcgacagctg	1767
QY	1572	tatcaacacagcgtcctctccgaccttccatcgcacaaagagcctcctctgagacagaa	1631
Db	1768	tatcaacacagcgtcctctccgaccttccatcgcacaaagagcctcctctgagacagaa	1827
QY	1632	tctttgcagaaagtgtgaaaaaacaccacttgcagaaatctctgtgaagcctataaacatcaat	1691
Db	1828	tctttgcagaaagtgtgaaaaaacaccacttgcagaaatctctgtgaagcctataaacatcaat	1887
QY	1692	tccctgtgatactgttgcatccattatatacagaagagatcacatccaaatagaacctgagc	1751
Db	1888	tccctgtgatactgttgcatccattatatacagaagagatcacatccaaatagaacctgagc	1947
QY	1752	caagaatttgaagcttctctccaaggtaaaagcttatataccaactcagaggaacatcccc	1811
Db	1948	caagaatttgaagcttctctccaaggtaaaagcttatataccaactcagaggaacatcccc	2007
QY	1812	gattactatttgcactctcttgaacaatttggccaatttggcgaagtgtctctgacttcaat	1871
Db	2008	gattactatttgcactctcttgaacaatttggccaatttggcgaagtgtctctgacttcaat	2067
QY	1872	aaactggaactttataggggagctatgctgtcaatgggaaaagcttcagagaacacagct	1931
Db	2068	aaactggaactttataggggagctatgctgtcaatgggaaaagcttcagagaacacagct	2127
QY	1932	ggaatccacatggaagaagcccccagaaacctatactcccgagagcgtgtatcttcttc	1991
Db	2128	ggaatccacatggaagaagcccccagaaacctatactcccgagagcgtgtatcttcttc	2187
QY	1992	tccaactgtgaagcaggaattcagagctctggaggtcacactccgggaatttcagcaagtgtg	2051
Db	2188	tccaactgtgaagcaggaattcagagctctggaggtcacactccgggaatttcagcaagtgtg	2247
QY	2052	aataagcaagatatacctatctcgtgggaaaataatctagcctctgcacaaagcctcagagctg	2111
Db	2248	aataagcaagatatacctatctcgtgggaaaataatctagcctctgcacaaagcctcagagctg	2307
QY	2112	caaatataagagatgtgctgtgtgtgtgtggaagcctcagtttggctccctcagcaccgtgtaag	2171
Db	2308	caaatataagagatgtgctgtgtgtgtgtgtggaagcctcagtttggctccctcagcaccgtgtaag	2367

QY	2172	aaacttattctcccaatgtgtggaagccagctccctccacacatagaagaatgaaagccatc	2231
Db	2368	aaacttattctcccaatgtgtggaagccagctccctccacacatagaagaatgaaagccatc	2427
QY	2232	acatctgtatacaaaccttgaaaaccttgatattcatgactacagaaatcaacagctcgg	2291
Db	2428	acatctgtatacaaaccttgaaaaccttgatattcatgactacagaaatcaacagctcgg	2487
QY	2292	gtgtgtctgactgacagcttgggtgaactttgagaacctataaagctcraatagatatac	2351
Db	2488	gtgtgtctgactgacagcttgggtgaactttgagaacctataaagctcraatagatatac	2547
QY	2352	ataaagaatgaatgaaagaaatgtcctataaactaagctgaagccctgaaaaaccttgaagaag	2411
Db	2548	ataaagaatgaatgaaagaaatgtcctataaactaagctgaagccctgaaaaaccttgaagaag	2607
QY	2412	atgtgttatttctcaatttggaccacttgtctgaactttgagaagggaaatgatatatactg	2471
Db	2608	atgtgttatttctcaatttggaccacttgtctgaactttgagaagggaaatgatatatactg	2667
QY	2472	aagctctgtcacaagtgaaccttggaccttgaaccttgaagaatctcaattagctctcctgcgtgtg	2531
Db	2668	aagctctgtcacaagtgaaccttggaccttgaaccttgaagaatctcaattagctctcctgcgtgtg	2727
QY	2532	tctgcaaatgtcagttgaaatctcctagctcagaatcttccaactttgttgcataactgagacat	2591
Db	2728	tctgcaaatgtcagttgaaatctcctagctcagaatcttccaactttgttgcataactgagacat	2787
QY	2592	cttgattattatcagaanaatttactcttgaaaaaagtgaatgaaatgtcattcattgaactgtac	2655
Db	2788	cttgattattatcagaanaatttactcttgaaaaaagtgaatgaaatgtcattcattgaactgtac	2847
QY	2652	gacagagatgaacgtgtctgagaacacgtcacacagctagatgtgcctctgggtgtgtgacgtg	2711
Db	2848	gacagagatgaacgtgtctgagaacacgtcacacagctagatgtgcctctgggtgtgtgacgtg	2907
QY	2712	caagcgagccttgagcagcctgttgaacaatttggagaagagtcaccaacactcgtcaagctt	2771
Db	2908	caagcgagccttgagcagcctgttgaacaatttggagaagagtcaccaacactcgtcaagctt	2967
QY	2772	gggttgaaaaaactggagactcacagatacagaattagaattcttagtgtgcatcttcttga	2831
Db	2968	gggttgaaaaaactggagactcacagatacagaattagaattcttagtgtgcatcttcttga	3027
QY	2832	aagaaacctcttgaaaaaacttccaagcagttgaattttggcgggaaaaactcgtgtgagcagtgat	2891
Db	3028	aagaaacctcttgaaaaaacttccaagcagttgaattttggcgggaaaaactcgtgtgagcagtgat	3087
QY	2892	ggatggtcttgccttcaaggggtgtattttgagaaattttaagcaattagtttttttggactt	2951
Db	3088	ggatggtcttgccttcaaggggtgtattttgagaaattttaagcaattagtttttttggactt	3147
QY	2952	agttacaagaatttctaacctgtacacagcatatgacaagaaccttagcagaagtgtatcc	3011
Db	3148	agttacaagaatttctaacctgtacacagcatatgacaagaaccttagcagaagtgtatcc	3207
QY	3012	aagttaaacttcttcgacaagaagctagagctgtgtgggtgcgaatttgatgatgatctc	3071
Db	3208	aagttaaacttcttcgacaagaagctagagctgtgtgggtgcgaatttgatgatgatctc	3267
QY	3072	agtttatttccaggtgtcttttaactagttaactgttaacttaataaagtgtaactcgaagcag	3131
Db	3268	agtttatttccaggtgtcttttaactagttaactgttaacttaataaagtgtaactcgaagcag	3327
QY	3132	ta 3133	11
Db	3328	ta 3329	11

RESULT	3
AAS03946	
ID	AAS03946 standard; DNA: 3615 BP.
XX	

AC	AAS03946,
XX	
DT	12-SEP-2001 (first entry)
XX	
DE	Human caspase recruitment domain 12 (CARD-12) genomic DNA.
XX	
KM	Caspase recruitment domain; CARD-12; apoptosis; stress-related pathway;
KM	cancer; viral infection; poxvirus; adenovirus; autoimmune disorder;
KW	systemic lupus erythematosus; arthritis; neurological disorder; stroke;
KM	Alzheimer's disease; amyotrophic lateral sclerosis; hematologic disease;
KV	aplastic anaemia; myocardial infarction; inflammatory disorder;
KM	Crohn's disease; insulin-dependent diabetes; contact dermatitis;
KM	psoriasis; graft rejection; bacterial infection; lepromatous leprosy;
KW	tuberculosis; ischemic brain injury; hypoxic brain injury; ds;
KM	kidney ischaemia; reperfusion injury; acute bacterial meningitis;
KX	excitotoxic brain damage; liver disease.
XX	
OS	Homo sapiens.
FH	
FT	
Key	Location/Qualifiers
CDS	1..3615
	/tag= a
	/product= "Human CARD-12"
PN	MO200130971-A2.
XX	
PD	03-MAY-2001.
XX	
PF	26-OCT-2000; 2000WO-US29643.
XX	
PR	27-OCT-1999; 99US-0161822.
XX	
PA	(MILL-) MILLENNIUM PHARM INC.
PI	Bertin J, Robison KE;
XX	
DR	WPI: 2001-308628/32.
XX	
XX	P-PSDB; AAU02881.
PT	Isolated caspase recruitment domain-12 polypeptide and nucleic acids
PT	encoding them, useful for treating and diagnosing disorders associated
PT	with abnormal apoptosis such as cancer, arthritis and Alzheimer's
PT	disease -
PS	
XX	
PS	Disclosure; Fig 2; 93pp; English.
CC	
CC	The sequence represents a genomic DNA which encodes the human caspase
CC	recruitment domain 12 (CARD-12) polypeptide. CARD domains are found in a
CC	number of proteins that transmit signals that activate apoptosis and
CC	Inflammatory pathways in response to stress and other stimuli. Therefore,
CC	CARD-12 and its corresponding nucleic acid may be used in treatment and
CC	diagnosis of patients suffering from disorders associated with an
CC	abnormal level (an increase or a decrease) of apoptotic cell death or
CC	abnormal activity of stress-related pathways. The disorders include
CC	cancer, viral infections (e.g. caused by poxviruses, adenoviruses),
CC	autoimmune disorders (e.g. systemic lupus erythematosus, arthritis),
CC	neurological disorders (e.g. Alzheimer's disease, amyotrophic lateral
CC	sclerosis), haematologic diseases (e.g. aplastic anaemia, myocardial
CC	infarction, stroke), inflammatory and immune system disorders (e.g.
CC	Crohn's disease, insulin-dependent diabetes, contact dermatitis,
CC	psoriasis, graft rejection), bacterial infections (e.g. tuberculosis,
CC	lepromatous leprosy), ischemic and hypoxic brain injury, kidney
CC	ischaemia/reperfusion injury, excitotoxic brain damage, acute bacterial
CC	meningitis and liver disease.
XX	
SO	Sequence 3615 BP; 1041 A; 811 C; 845 G; 918 T; 0 other;

Query Match 84.1% Score 2634; DB 22; Length 3615;
 Best Local Similarity 99.9%; Pred No. 0;
 Matches 2784; Conservative 0; Mismatches 3; Indels 0; Gaps 0

Db	734	gctctttccatcagacatcaagaagactgagactgttgctcagagatttaaaagact	793
Qy	358	tgtaaccatcccccaactctttctgaactttatcccccttggtagagatatgacattat	417
Db	794	tgtaaccatacccccatctttcttgaaactttatcccccttggtagagatatgacattat	853
Qy	418	ttcaacttgaanaagcactcttcacagaacctgtctctgtgagaagaagacacacacac	477
Db	854	ttcaacttgaanaagcactcttcacagacctgtctctgtgagaagaagacacacacac	913
Qy	478	gcgtgtgagcagctgacccctgaaatggctctctgcaagctcttcagacctgcatatg	537
Db	914	gcgtgtgagcagctgacccctgaaatggctctctctcagagctcttcagacctgcatatg	973
Qy	538	aaggggaactctgacaaagcgaafcccaactctgtctgcagcgacattgcaagctctgggt	597
Db	974	aaggggaactctgacaaagcgaafcccaactctgtctgcagcgacattgcaagctctgggt	1033
Qy	598	ccggaaagtgcgaagctctgcacaaagtccaatctgtctctctccctccgtctccagcag	657
Db	1034	ccggaaagtgcgaagctctgcacaaagtccaatctgtctctctccctccgtctccagcag	1099
Qy	658	cccaggtgtgacattttgnaaacctctgtgtatcaactctctgatatccctgacaaatca	717
Db	1094	cccaggtgtgacattttgnaaacctctgtgtatcaactctctgatatccctgacaaatca	1153
Qy	718	ggaagcagaacatctcaatggcccatgtctgtgaagctggcgagagaggtctctctctgt	777
Db	1154	ggaagcagaacatctcaatggcccatgtctgtgaagctggcgagagaggtctctctctgt	1211
Qy	778	atggtctacaatgatctcaagccccaagaactgtgccagaatatcgaaagccctgataaagaa	837
Db	1214	atggtctacaatgatctcaagccccaagaactgtgccagaatatcgaaagccctgataaagaa	1272
Qy	838	accacccgcttcaagaacatggtcatcgttcacacactacactgtgactgtaggcacatac	897
Db	1274	accacccgcttcaagaacatggtcatcgttcacacactacactgtgactgtaggcacatac	1333
Qy	898	ggcgaatttgggtcccttgactctctgaaggctgggggatatgagaagaagccagccgaagctc	957
Db	1334	ggcgaatttgggtcccttgactctctgaaggctgggggatatgagaagaagccagccgaagctc	1393
Qy	958	tcataccgaagaagtgcgtatcaagaagagctgtctgaagagctgtgtgctccaaattcsgaaat	1017
Db	1394	tcataccgaagaagtgcgtatcaagaagagctgtctgaagagctgtgtgctccaaattcsgaaat	1455
Qy	1018	ccaggtgtcttgaagaaatctcatgaagacccctctcttvtgtatcactgtgtgcaatcc	1077
Db	1454	ccaggtgtcttgaagaaatctcatgaagacccctctcttvtgtatcactgtgtgcaatcc	1514
Qy	1078	agaatgggtgaagaagtgtgaattcaactctcaacacaaacaaagcgtgttcataactctctatg	1137
Db	1514	agaatgggtgaagaagtgtgaattcaactctcaacacaaacaaagcgtgttcataactctctatg	1573
Qy	1138	atctgttcatatcacgaanaaacaacaacaaatagaagtgctgtctgcaagtgtgactcatcc	1197
Db	1574	atctgttcatatcacgaanaaacaacaacaaatagaagtgctgtctgcaagtgtgactcatcc	1633
Qy	1198	ggagccctgagaccactgtgtggaacctagctcttvgagaggctgtgtctcccaagaattgatac	1257
Db	1634	ggagccctgagaccactgtgtggaacctagctcttvgagagggtgtgtctcccaagaattgatac	1693
Qy	1258	tcgaactctgcaagatgtgtccgcggttgaaatgaagatgtctgcctgcgaanaacttggccctct	1314
Db	1694	tcgaactctgcaagatgtgtccgcggttgaaatgaagatgtctgcctgcgaanaacttggccctct	1755
Qy	1318	gtaaatatacagctccaagaagtgttcaagccgaagatataattcttccaagtcataccag	1377
Db	1754	gtaaatatacagctccaagaagtgttcaagccgaagatataattcttccaagtcataccag	1813
Qy	1378	agtaacagcaggaacgaagaatcagcaggtttatgtacgttccatgtgccaagagaggtgta	1437

Db	1814	agtacacagcaggaacgaacccacagcagttatctgacgtcctcagccagagcagsggtga	1873
Qy	1438	ccaagggagatggttactctgcagaaabtggttcacatttcgcagacatccatccattata	1497
Db	1874	ccaagggagatggttactctgcagaaabtggttcacatttcgcagacatccatccattata	1933
Qy	1498	gcagcctgtctccggttaacactctgtgtgtcactctgtgtgaagccacagsgctgttatgaagc	1557
Db	1934	gcagcctgtctccggttaacactctgtgtgtcactctgtgtgaagccacagsgctgttatgaagc	1993
Qy	1558	acctcccaacgtgtatatacaacgctgcctctctcgcagacttccatctgcagagagccctc	1617
Db	1994	acctcccaacgtgtatatacaacgctgcctctctcgcagacttccatctgcagagagccctc	2053
Qy	1618	tcctggagacaggaatctcttgcacaaagtgtgaaacaacacacactgagcaagaatctcgaaag	1677
Db	2054	tcctggagacaggaatctcttgcacaaagtgtgaaacaacacacactgagcaagaatctcgaaag	2113
Qy	1678	ccataaacaatcaatccctcttctgtgaggtgtgtgcacatccattatatacaagagttacatcca	1737
Db	2114	ccataaacaatcaatccctcttctgtgaggtgtgtgcacatccattatatacaagagttacatcca	2173
Qy	1738	aatcgcccttgagccaagaatttgaaagcttctcttccaagtgtaaaagcttatataccaact	1797
Db	2174	aatcgcccttgagccaagaatttgaaagcttctcttccaagtgtaaaagcttatataccaact	2233
Qy	1798	cagggaacatcccccagatttacttacttgactctcttgcacattctgcacattgtgcgaatgt	1857
Db	2234	cagggaacatcccccagatttacttacttgactctcttgcacattctgcacattgtgcgaatgt	2293
Qy	1858	ctctggaacttcaatcaactggaacttltatgtgggagctatgtgctcatgtggaanaagsgctg	1917
Db	2294	ccctggaacttcaatcaactggaacttltatgtgggagctatgtgctcatgtggaanaagsgctg	2353
Qy	1918	cagaagacacaggtgtgaatccacatgtagaagagcccccagaaacccacatctccagcagag	1977
Db	2354	cagaagacacaggtgtgaatccacatgtagaagagcccccagaaacccacatctccagcagag	2413
Qy	1978	ctgtactcttctgtcttccaacttggaagcaggaattcagagctctgtgaagttcaaccccgag	2037
Db	2414	ctgtactcttctgtcttccaacttggaagcaggaattcagagctctgtgaagttcaaccccgag	2473
Qy	2038	atttcagcaagtttgataaagaacagatatacaatctgtggggaanaattatcagctctgcca	2097
Db	2474	atttcagcaagtttgataaagaacagatatacaatctgtggggaanaattatcagctctgcca	2533
Qy	2098	caagcctcgaagctgtgcaataaagaagatgtgctggtgtgtgtcgtggaagcctcagttgtgcc	2157
Db	2534	caagcctcgaagctgtgcaataaagaagatgtgctggtgtgtgtcgtggaagcctcagttgtgcc	2593
Qy	2158	tcagcacactgtataagacatttattctctccatgtgtggaagcagctccctccacatagagag	2217
Db	2594	tcagcacactgtataagacatttattctctccatgtgtggaagcagctccctccacatagagag	2653
Qy	2218	atgagagggacacatcacatctgttaacaaacctgaaaccttgagttatcatgagactacaga	2277
Db	2654	atgagagggacacatcacatctgttaacaaacctgaaaccttgagttatcatgagactacaga	2713
Qy	2278	atcaacgctgtccgggtgtgtctgcagctgacagcttgggttaacttgagaacacttacaaagc	2337
Db	2714	atcaacgctgtccgggtgtgtctgcagctgacagcttgggttaacttgagaacacttacaaagc	2773
Qy	2338	tcatatgataaacaataagatgtagaagaagatgctataaanaactagaatgaagccctga	2397
Db	2774	tcatatgataaacaataagatgtagaagaagatgctataaanaactagaatgaagccctga	2833
Qy	2398	aaaacctgaagaagaatgtgttatttctatctgtgacccacttgtctgaactctggaagggaa	2457
Db	2834	aaaacctgaagaagaatgtgttatttctatctgtgacccacttgtctgaactctggaagggaa	2893
Qy	2458	tgtatatactatgataaggtctctgtctcaagtgaaacctgtgacacttgaagaanaattcaatag	2517
Db	2894	tgtatatactatgataaggtctctgtctcaagtgaaacctgtgacacttgaagaanaattcaatag	2953

Qy	2518	tctcctgtctgtctgtcgtcaaatgcagtgaaatctcctagctccagaaatcttcacatttgg	2577
Db	2954	tctcctgtctgtctgtcgtcaaatgcagtgaaatctcctagctccagaaatcttcacatttgg	3013
Qy	2578	tcaactctgacatctctgtattatccagaanaattacactctggaanaagaatgaaatgaagctc	2637
Db	3014	tcaactctgacatctctgtattatccagaanaattacactctggaanaagaatgaaatgaagctc	3073
Qy	2638	ttcatgaactctgatacagcaggaatgaacgtgtcagagaagcctcagcagactgtgtgcctc	2697
Db	3074	ttcatgaactctgatacagcaggaatgaacgtgtcagagaagcctcagcagactgtgtgcctc	3133
Qy	2698	ggggctgtgacgtgtgcaagcagcctgtgagcagcctgttgaacatttggagaggtcccaac	2757
Db	3134	ggggctgtgacgtgtgcaagcagcctgtgagcagcctgttgaacatttggagaggtcccaac	3193
Qy	2758	aactcgtcaagcttgggtgtgtaaaaaactgtgagactccacagatatacagattgaatttag	2817
Db	3194	aactcgtcaagcttgggtgtgtaaaaaactgtgagactccacagatatacagattgaatttag	3253
Qy	2818	gtgcattttttggaaagacccctctgnaaaaaactccagcagttgaatttggcgggaatc	2877
Db	3254	gtgcattttttggaaagacccctctgnaaaaaactccagcagttgaatttggcgggaatc	3313
Qy	2878	gtgtgagcagtgatgagatgagctgtccctcatcaggtgtgatttggaatcttgaacattag	2937
Db	3314	gtgtgagcagtgatgagatgagctgtccctcatcaggtgtgatttggaatcttgaacattag	3373
Qy	2938	tgtttttgactttagtactaaagaattctactacatgataccagatagtcagaaactta	2997
Db	3374	tgtttttgactttagtactaaagaattctactacatgataccagatagtcagaaactta	3433
Qy	2998	gccaaagtgtatccaaagttaactttctcgcagaagaagcttagctgtgtgtgtgcgaatttg	3057
Db	3434	gccaaagtgtatccaaagttaactttctcgcagaagaagcttagctgtgtgtgtgcgaatttg	3493
Qy	3058	atgatgatgcactcagttgatttcaag	3084
Db	3494	atgatgatgcactcagttgatttcaag	3520

RESULT	4
AAH99581	
ID	AAH99581 standard; cDNA; 2950 BP.
XX	
AC	AAH99581;
XX	
DT	16-OCN-2001 (first entry)
XX	
DE	Human protein encoding cDNA sequence SEQ ID NO:416.
XX	
KW	Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
KW	antiInflammatory; antiInflammatory; antiarthritic; immunosuppressive;
KW	antibacterial; endocrine; cardient; central nervous system; virulence;
KW	anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
KW	antiagregant; haemostatic; vulnerability; antiulcer; osteopathic; eczema;
KW	dermatological; antiallergic; antiasthmatic; antidiabetic; cyrostatic;
KW	neuroprotective; antidepressant; nootropic; antiParkinsonian; infection;
KW	immunostimulant; gene therapy

XX 22-DEC-2000; 2000WO-US35017.
XX
XX 23-DEC-1999; 99US-0471275.
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
XX
XX (HXSE-) HXSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT;
XX
XX WPI: 2001-457603/49.
DR P-PSDB: AAM25640.
XX
XX Isolated human polynucleotides encoding polypeptides, useful for the
PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
XX
XX Claim 1; Page 511-512; 1217pp; English.
XX
XX AAH9166 to AAH99904 encode the human proteins given in AAM25225 to
CC AAM25963. The proteins can have activities based on the tissues and
CC cells they are expressed in, such as: antiinflammatory; antirheumatic;
CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac;
CC central nervous system; virocid; anti-HIV; fungicide; antimutagen;
CC cardiovascular; antianemic; antiagregant; haemostatic; vulnery;
CC antiulcer; osteopathic; dermatological; antiallergic; antiaesthetic;
CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
CC encoding them can be used in gene therapy, antisense therapy and vaccine
CC production. The proteins and polynucleotides are useful for screening for
CC agonists or antagonists of a protein and for the treatment and diagnosis
CC of disorders associated with the activity of a protein e.g. inflammation,
CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC neuroendology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
CC osteoporosis, severe combined immunodeficiency, eczema, allergic
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC neurological disorders.
XX
XX Sequence 2950 BP; 915 A; 592 C; 628 G; 815 T; 0 other:
SO
Query Match 57.7%; Score 1807; DB 22; Length 2950;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1907; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1225 ctctggaggggtgtgttcccccaagaattgttgcgaactgcagatgttgcgcagctga 1284
b 45 ctctggaggggtgtgttcccccaagaattgttgcgaactgcagatgttgcgcagctga 104
QY 1285 atgaggaatgtctctgcgaactgcgcctcctgttaataatacaagcttaaaagttcaagc 1344
Db 105 atgaggaatgtctctgcgaactgcgcctcctgttaataatacaagcttaaaagttcaagc 164
QY 1345 caaagataaattcttcaagaatcattcagaggtacacagcagcaagaactcagca 224
Db 165 caaagataaattcttcaagaatcattcagaggtacacagcagcaagaactcagca 224
QY 1405 gtttatatagcttcataagcagagaggtgacccaaggggaatggttacttgcagaana 1464
Db 225 gtttatatagcttcataagcagagaggtgacccaaggggaatggttacttgcagaana 284
QY 1465 tggttccatttcgacattacacacttataagcagctctccgcgtacacactgtggt 1524
Db 285 tggttccatttcgacattacacacttataagcagctctccgcgtacacactgtggt 344
QY 1525 catctgtggaagcacaagggctgttatgaagacactgcagcagtgatatcaacagcgt 1584
Db 345 catctgtggaagcacaagggctgttatgaagacactgcagcagtgatatcaacagcgt 404
QY 1585 guctctcgcgacttcacatgcgaagagcctctctgtgagacgaagaacttgcagaagt 1644

Db 405 guctctcgcgacttcacatgcgaagagcctctctgtgagacgaagaacttgcagaagt 464
QY 1645 tgaanaaacccactgagcaagaatcttgcgaagacataaacatcattcctgttagagt 1704
Db 465 tgaanaaacccactgagcaagaatcttgcgaagacataaacatcattcctgttagagt 524
QY 1705 gtgcatacattatatacaagaggtacataccaatcagccctgcagcagaagatttgaag 1764
Db 525 gtgcatacattatatacaagaggtacataccaatcagccctgcagcagaagatttgaag 584
QY 1765 cttcttccaaggtataaagcttatatacactcaggggaacatcccccattactatttg 1824
Db 585 cttcttccaaggtataaagcttatatacactcaggggaacatcccccattactatttg 644
QY 1825 acttcttgaacatttgcgaacttgcgaagtgcctgcgaacttcaactgaacgagccttt 1884
Db 645 acttcttgaacatttgcgaacttgcgaagtgcctgcgaacttcaactgaacgagccttt 704
QY 1885 atggggagctatgtgcttcattggaagaagctgcagaagacacaggttgaatccacatg 1944
Db 705 atggggagctatgtgcttcattggaagaagctgcagaagacacaggttgaatccacatg 764
QY 1945 aagagggcccgaaaacctacattcccaagggcgttatcttcttccaactggaagc 2004
Db 765 aagagggcccgaaaacctacattcccaagggcgttatcttcttccaactggaagc 824
QY 2005 aggaattcagaactctggaaggtcacaccccggaattcagaagttgataagaagaata 2064
Db 825 aggaattcagaactctggaaggtcacaccccggaattcagaagttgataagaagaata 884
QY 2065 tcaatattcggggaanaatatcagctctgcacaaagcctcagagctgcanaataaagagt 2124
Db 885 tcaatattcggggaanaatatcagctctgcacaaagcctcagagctgcanaataaagagt 944
QY 2125 gtgcgggtgtgtgtggaagcctcagttgtgtcctaagaacctgtgaagaacttattctc 2184
Db 945 gtgcgggtgtgtgtggaagcctcagttgtgtcctaagaacctgtgaagaacttattctc 1004
QY 2185 tcatgtgtggaagccagctccctcacacataagaagatggaagacatacattgttaacaa 2244
Db 1005 tcatgtgtggaagccagctccctcacacataagaagatggaagacatacattgttaacaa 1064
QY 2245 acctgaaaaccttgaattcattacagcactcagaaatcaacagcgtgcgcgggtgctcagtg 2304
Db 1065 acctgaaaaccttgaattcattacagcactcagaaatcaacagcgtgcgcgggtgctcagtg 1124
QY 2305 acagcttgggttaacttgaagaaccttacaagctcattatgatacagataaagaatgaatg 2364
Db 1125 acagcttgggttaacttgaagaaccttacaagctcattatgatacagataaagaatgaatg 1184
QY 2365 aagaagatgctataaatactgaagcctgcgaagaacctgcgaagaagatgtgttattctc 2424
Db 1185 aagaagatgctataaatactgaagcctgcgaagaacctgcgaagaagatgtgttattctc 1244
QY 2425 atttgaccactgtctgcacatttgaagaggaatggaattacatagctcaagtcctgtcaa 2484
Db 1245 atttgaccactgtctgcacatttgaagaggaatggaattacatagctcaagtcctgtcaa 1304
QY 2485 gttgaacctgtgaccttgaagaataatcattagcttctcgtgcgtgtgtcgaatgtag 2544
Db 1305 gttgaacctgtgaccttgaagaataatcattagcttctcgtgcgtgtgtcgaatgtag 1364
QY 2545 tgaatactcagctcagaatcttcaacaatttgcgttaaaactgagcaattcttattatcag 2604
Db 1365 tgaatactcagctcagaatcttcaacaatttgcgttaaaactgagcaattcttattatcag 1424
QY 2605 aaatctacgtggaagaagaatggaagaatggaagcttcatgaaactgaaatcagaatgaaag 2664
Db 1425 aaatctacgtggaagaagaatggaagaatggaagcttcatgaaactgaaatcagaatgaaag 1484
QY 2665 tgcctagaacacgtcaccgacattgtagctgcctcgggtgtgtgagctgtgaagggcagcctga 2724

Db	1485	tgctagaagaagctccacgcacctgtagctgctcgccggtgcgtgcagtgcaagcgagccctga	1544
QY	2725	gcagccctgtgtgaacaacatttggagaggtccccaacactcgtcaagcttggtgtgaaaaact	2784
Db	1545	gcagccctgtgtgaacaacatttggagaggtccccaacactcgtcaagcttggtgtgaaaaact	1604
QY	2785	gggaactcacagabacagagatagaattttagtgtagctatttttggaaagaacctctga	2844
Db	1605	gggaactcacagabacagagatagaattttagtgtagctatttttggaaagaacctctga	1664
QY	2845	aaaactccagcagttgaatttggcgggaatcgtgtgacgaatgaatggatggtctgct	2904
Db	1665	aaaactccagcagttgaatttggcgggaatcgtgtgacgaatgaatggatggtctgct	1724
QY	2905	tcatgtgtgtcatatttgagaatcctaagccaattagtgctttttgaccttgtaactaaagaat	2964
Db	1725	tcatgtgtgtcatatttgagaatcctaagccaattagtgctttttgaccttgtaactaaagaat	1784
Y	2965	ttctacctgatccaagcattatgccaagaanaacttaagccaagtgttataccaagttaactttc	3024
Db	1785	ttctacctgatccaagcattatgccaagaanaacttaagccaagtgttataccaagttaactttc	1844
QY	3025	tgcagaagcctagtcctgtgtgggtggtcattttgagatgatatctcaagtgtttatcaag	3084
Db	1845	tgcagaagcctagtcctgtgtgggtggtcattttgagatgatatctcaagtgtttatcaag	1904
QY	3085	gtgtgttttaactagtaactgccttaataaagtgtatctccgaagccagta	3133
Db	1905	gtgtgttttaactagtaactgccttaataaagtgtatctccgaagccagta	1953

CC expressed with decreased expression by rectifying mutations or deletions
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patient's own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAH77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ Sequence 2735 BP; 791 A; 555 C; 512 G; 876 T; 1 other;

```

KW cervical cancer; ss.
XX
XX Homo sapiens.
XX
XX WO200157278-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00670.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488901/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human cervical epithelial cells -
XX
XX Claim 25; SEQ ID NO 4322; 487bp; English.
XX
XX The present invention relates to human single exon nucleic acid probes
XX (SENPs). The present sequence is one such probe. The SENPs are derived
XX from human HeLa cells. The SENPs can be used to produce a single exon
XX microarray, which can be used for measuring human gene expression in a
XX sample derived from human cervical epithelial cells. By measuring gene
XX expression, the probes are therefore useful in grading and/or staging
XX of diseases of the cervix, notably cervical cancer.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pat_sequences.
XX
XX Sequence 421 BP; 126 A; 69 C; 77 G; 149 T; 0 other;
XX
XX
XX Query Match 7.7%; Score 242; DB 22; Length 421;
XX Best Local Similarity 100.0%; Pred. No. 7.6e-110;
XX Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 2816 aggtgcatcttttggaaagaccctctgaaacctccagcagttgaattggcgggaaa 2875
XX |||||||
XX Db 180 aggtgcatcttttggaaagaccctctgaaacctccagcagttgaattggcgggaaa 239
XX |||||||
XX 2876 tcgtgtgacgagtgatgagtgctgcctcatgggtgtatttggaaacttaagaact 2935
XX |||||||
XX Db 240 tcgtgtgacgagtgatgagtgctgcctcatgggtgtatttggaaacttaagaact 299
XX |||||||
XX 2936 agtgttttttactttagtactaaagaattcttacctgataccagattagtcagaanaact 2995
XX |||||||
XX Db 300 agtgttttttactttagtactaaagaattcttacctgataccagattagtcagaanaact 359
XX |||||||
XX 2996 tagccaagtgtatccaagttaactttctgcagaagctagagctgtgtgggtgcaatt 3055
XX |||||||
XX Db 360 tagccaagtgtatccaagttaactttctgcagaagctagagctgtgtgggtgcaatt 419
XX |||||||
XX 3056 tg 3057
XX ||
XX Db 420 tg 421
XX
XX RESULT 7
XX AAI35764
XX ID AAI35764 standard; DNA; 421 BP.
XX
XX AC, AAI35764;
```

```

XX
XX 17-OCT-2001 (first entry)
XX
XX Probe #4450 used to measure gene expression in human placenta sample.
XX
XX DE Probe; microarray; human; placenta; antenatal diagnosis;
XX KW genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX WO200157272-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00663.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488897/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX
XX Claim 25; SEQ ID NO 4450; 654bp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENPs).
XX The present sequence is one such probe. The probes are useful for
XX producing a microarray for predicting, measuring and displaying gene
XX expression in samples derived from human placenta. The probes are useful
XX for antenatal diagnosis of human genetic disorders.
XX
XX Sequence 421 BP; 126 A; 69 C; 77 G; 149 T; 0 other;
XX
XX
XX Query Match 7.7%; Score 242; DB 22; Length 421;
XX Best Local Similarity 100.0%; Pred. No. 7.6e-110;
XX Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 2816 aggtgcatcttttggaaagaccctctgaaacctccagcagttgaattggcgggaaa 2875
XX |||||||
XX Db 180 aggtgcatcttttggaaagaccctctgaaacctccagcagttgaattggcgggaaa 239
XX |||||||
XX 2876 tcgtgtgacgagtgatgagtgctgcctcatgggtgtatttggaaacttaagaact 2935
XX |||||||
XX Db 240 tcgtgtgacgagtgatgagtgctgcctcatgggtgtatttggaaacttaagaact 299
XX |||||||
XX 2936 agtgttttttactttagtactaaagaattcttacctgataccagattagtcagaanaact 2995
XX |||||||
XX Db 300 agtgttttttactttagtactaaagaattcttacctgataccagattagtcagaanaact 359
XX |||||||
XX 2996 tagccaagtgtatccaagttaactttctgcagaagctagagctgtgtgggtgcaatt 3055
XX |||||||
XX Db 360 tagccaagtgtatccaagttaactttctgcagaagctagagctgtgtgggtgcaatt 419
XX |||||||
XX 3056 tg 3057
XX ||
XX Db 420 tg 421
XX
XX RESULT 8
XX AAI04213
XX ID AAI04213 standard; DNA; 421 BP.
XX
XX
```

AC AAI04213;
XX
DT 09-OCT-2001 (first entry)
XX
DE Probe #4204 used to measure gene expression in human breast sample.
XX
KW Probe: human; breast disease; breast cancer; development disorder; ss;
KM Inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
OS Homo sapiens.
XX
PN WO200157270-A2.
XX
PD 09-AUG-2001.
XX
PF 29-JAN-2001; 2001WO-US00661.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-476286/51.
XX
PT Novel single exon nucleic acid probe used to measuring gene expression
PT in a human breast -
XX
PS Claim 25; SEQ ID NO 4204; 322pp; English.
XX
CC The present invention relates to novel single exon nucleic acid probes.
CC The present sequence is one such probe. The probes are useful for
CC measuring human gene expression in a human breast sample, where the probe
CC hybridises at high stringency to a nucleic acid expressed in the human
CC breast. The probes are useful for predicting, diagnosing, grading,
CC staging, monitoring and prognosing diseases of the human breast,
CC particularly those diseases with polygenic aetiology. The diseases
CC include: breast cancer, disorders of development, inflammatory diseases
CC of the breast, fibrocystic changes, proliferative breast disease and
CC non-carcinoma tumours.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 421 BP; 126 A; 69 C; 77 G; 149 T; 0 other;

Query Match 7.7%; Score 242; DB 22; Length 421;
Best Local Similarity 100.0%; Pred. No. 7.6e-110;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2816 aggtgcatttttggaaagaacctctgaaatactccagcagttgaatttgcgggaa 2875
DB |||||||
DB 180 aggtgcatttttggaaagaacctctgaaatactccagcagttgaatttgcgggaa 239
QY 2876 tctgtgagcagtgatgagctgcctcatggtggtatattagaatttaagaatt 2935
DB |||||||
DB 240 tctgtgagcagtgatgagctgcctcatggtggtatattagaatttaagaatt 299
QY 2936 agtgcatttttgaacttgaactaaagaatttctaccctgaccagcattagtcagaaact 2995
DB |||||||
DB 300 agtgcatttttgaacttgaactaaagaatttctaccctgaccagcattagtcagaaact 359
QY 2996 taagcagtgattatccagtaactttctgcagaagaagcagctgttgggtggcaat 3055
DB |||||||
DB 360 taagcagtgattatccagtaactttctgcagaagaagcagctgttgggtggcaat 419

QY 3056 tg 3057
DB ||
DB 420 tg 421

RESULT 9
AAI23590
ID AAI23590 standard; DNA: 220 BP.
XX
AC AAI23590;
XX
DT 12-OCT-2001 (first entry)
XX
DE Probe #13523 for gene expression analysis in human cervical cell sample.
XX
KW Probe: human; microarray; gene expression; cervical epithelial cell;
XX cervical cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157278-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00670.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX
PS Claim 25; SEQ ID NO 13523; 487pp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENPs). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 220 BP; 58 A; 32 C; 54 G; 76 T; 0 other;

Query Match 7.0%; Score 220; DB 22; Length 220;
Best Local Similarity 100.0%; Pred. No. 6.2e-99;
Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2865 ttgcgggaaatcggtgagcagtgatgagctgctcctcatggtggtataggaaat 2924
DB |||||||
DB 1 ttgcgggaaatcggtgagcagtgatgagctgctcctcatggtggtataggaaat 60
QY 2925 cttagaagaattaggttttgaacttgaactaaagaatttctaccctgaccagcatt 2984
DB |||||||
DB 61 cttagaagaattaggttttgaacttgaactaaagaatttctaccctgaccagcatt 120
QY 2985 gtccagaagaacttgaagtgattatccagtaactttctgcagaagaagcagctgtt 3044
DB |||||||

Db 121 gtcagaaacttagcgaagtgcttaccagaagttaactttctgcgaagaagctagcttgct 180
QY 3045 ggtgtagcaattgatgatgatgatcagtgcttattacag 3084
|||||
Db 181 ggtgtagcaattgatgatgatgatcagtgcttattacag 220

RESULT 10
ID AAI48904 standard; DNA; 220 BP.
XX AAI48904:
AC
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #17590 used to measure gene expression in human placenta sample.
XX
KW Probe: microarray; human; placenta; antenatal diagnosis;
KM genetic disorder; ss.
XX
OS Homo sapiens.
XX
XX W0200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
XX
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
DR WPI; 2001-488897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
PT

Claim 25; SEQ ID No 17590; 654pp; English.
XX
PS
XX The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
XX Sequence 220 BP; 58 A; 32 C; 54 G; 76 T; 0 other;
SQ

Query Match 7.0%; Score 220; DB 22; Length 220;
Best Local Similarity 100.0%; Pred. No. 6.2e-99;
Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2865 ttggcgggaaatcgtgtgagcagtgatgagtgcttgccttcacaggggtatttgagagat 2924
|||||
Db 1 ttggcgggaaatcgtgtgagcagtgatgagtgcttgccttcacaggggtatttgagagat 60

QY 2925 cttagcaattagtgcttttgcacttagtactaaagaatttaccctgacccaagcatca 2984
|||||
Db 61 cttagcaattagtgcttttgcacttagtactaaagaatttaccctgacccaagcatca 120

QY 2985 gtcagaaaaacttagccaagtgttaccagaagtttaactttctgcgaagaagctaggttgct 3044
|||||
Db 121 gtcagaaaaacttagccaagtgttaccagaagtttaactttctgcgaagaagctaggttgct 180

QY 3045 ggtgtagcaattgatgatgatgatcagtgcttattacag 3084

Db 181 ggtgtagcaattgatgatgatgatcagtgcttattacag 220
|||||

RESULT 11
ID AAI09206 standard; DNA; 220 BP.
XX AAI09206:
AC
XX
DT 09-OCT-2001 (first entry)
XX
DE Probe #9197 used to measure gene expression in human breast sample.
XX
KW Probe: human; breast disease; breast cancer; development disorder; ss;
KM inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
OS Homo sapiens.
XX
XX W0200157270-A2.
XX
PD 09-AUG-2001.
XX
PF 29-JAN-2001; 2001WO-US00661.
XX
PR 04-FEB-2000; 2000US-0180312.
XX
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
DR WPI; 2001-476286/51.
XX
PT Novel single exon nucleic acid probe used to measuring gene expression
XX in a human breast -
PT

Claim 25; SEQ ID No 9197; 322pp; English.
XX
PS
XX The present invention relates to novel single exon nucleic acid probes.
CC The present sequence is one such probe. The probes are useful for
CC measuring human gene expression in a human breast sample, where the probe
CC hybridises at high stringency to a nucleic acid expressed in the human
CC breast. The probes are useful for predicting, diagnosing, grading,
CC staging, monitoring and prognosing diseases of the human breast,
CC particularly those diseases with polygenic aetiology. The diseases
CC include: breast cancer, disorders of development, inflammatory diseases
CC of the breast, fibrocystic changes, proliferative breast disease and
CC non-carcinoma tumours.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 220 BP; 58 A; 32 C; 54 G; 76 T; 0 other;
SQ

Query Match 7.0%; Score 220; DB 22; Length 220;
Best Local Similarity 100.0%; Pred. No. 6.2e-99;
Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2865 ttggcgggaaatcgtgtgagcagtgatgagtgcttgccttcacaggggtatttgagagat 2924
|||||
Db 1 ttggcgggaaatcgtgtgagcagtgatgagtgcttgccttcacaggggtatttgagagat 60

QY 2925 cttagcaattagtgcttttgcacttagtactaaagaatttaccctgacccaagcatca 2984
|||||
Db 61 cttagcaattagtgcttttgcacttagtactaaagaatttaccctgacccaagcatca 120

OY 2985 gtcagaaactagccaagtgatccaaagttctctgcaagaagctagctgtt 3044
 |||||||
 Db 121 gtcagaaactagccaagtgatccaaagttctctgcaagaagctagctgtt 180
 OY 3045 gggctggcaattgatgatgatcattcatttattacag 3084
 |||||||
 Db 181 gggctggcaattgatgatgatcattcatttattacag 220

RESULT 12

AAH11452
 ID AAH11452 standard; cDNA; 549 BP.

AC AAH11452;

DT 26-JUN-2001 (first entry)

DE Human cDNA clone (3'-primer) SEQ ID NO: 8287.

KM Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.

OS Homo sapiens.

PN EP1074617-A2.

PD 07-FEB-2001.

PF 28-JUL-2000; 2000EP-0116126.

PR 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

PA (HELI-) HELIX RES INST.

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Makamatsu A, Nagai K, Otsuki T;

DR WPI: 2001-318749/34.

PT primer sets for synthesizing polynucleotides, particularly the 5602

PT full-length cDNAs defined in the specification, and for the detection

PT and/or diagnosis of the abnormality of the proteins encoded by the

PT full-length cDNAs -

PS Claim 3; SEQ ID 8287; 2537pp + CD ROM; English.

CC The present invention describes primer sets for synthesizing 5602

CC full-length cDNAs defined in the specification. Where a primer set

CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary

CC to the complementary strand of a polynucleotide which comprises one of

CC the 5602 nucleotide sequences defined in the specification, where the

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination

CC of an oligonucleotide comprising a sequence complementary to the

CC complementary strand of a polynucleotide which comprises a 5'-end

CC sequence and an oligonucleotide comprising a sequence complementary to a

CC polynucleotide which comprises a 3'-end sequence, where the

CC oligonucleotide comprises at least 15 nucleotides and the combination of

CC the 5'-end sequence/3'-end sequence is selected from those defined in

CC the specification. The primer sets can be used in antisense therapy and

CC in gene therapy. The primers are useful for synthesizing polynucleotides,

CC particularly full-length cDNAs. The primers are also useful for the

CC detection and/or diagnosis of the abnormality of the proteins encoded by

CC the full-length cDNAs. The primers allow obtaining of the full-length

CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and

CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to

CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632

CC represent oligonucleotides, all of which are used in the exemplification

CC of the present invention.

XX Sequence 549 BP; 162 A; 120 C; 119 G; 139 T; 9 other;

Query Match 0.6%; Score 19; DB 22; Length 549;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 72 caaagaatgggaatgactg 90
 |||||||
 Db 214 caaagaatgggaatgactg 232

RESULT 13

AAH30628/c
 ID AAH30628 standard; DNA; 579 BP.

AC AAH30628;

DT 08-JUN-1999 (first entry)

DE H. pylori cell envelope protein ORF hp7el0433_5345837_c2_8.

KW Vaccine; probe; diagnostic; ORF; cell envelope protein;

KW secreted protein; cytoplasmic protein; cellular protein; ds.

OS Helicobacter pylori.

PN WO9824475-A1.

PD 11-JUN-1998.

PF 05-DEC-1997; 97WO-US22104.

PR 14-JUL-1997; 97US-0891928.

PR 05-DEC-1996; 96US-0759625.

PR 25-MAR-1997; 97US-0823745.

PA (ASTR) ASTRA AB.

PI Alm RA, Castriotta LM, Dolg PC, Kabok Z, Smith D;

DR WPI: 1998-333051/29.

PT New isolated Helicobacter pylori nucleic acids - used to develop

PT products for the diagnosis, prevention and treatment of infection by

PT H. pylori and other Helicobacter species

PS Claims 3, 4; Page 169; 339pp; English.

CC Recombinant or substantially pure preparations of H. pylori polypeptides

CC are disclosed, together with the nucleic acids encoding them. In all,

CC 97 ORFs are shown. The proteins are variously cell envelope proteins,

CC cytoplasmic proteins, secreted proteins or other cellular proteins.

CC Vaccines containing the nucleic acids or proteins are claimed, as are

CC probes containing at least 8 nucleotides from the nucleic acid

CC sequences. The vaccines are useful for treating or reducing the risk of

CC H. pylori infections, and the probes can be used diagnostically for

CC detecting the presence of Helicobacter in a sample. The products are

CC also of use in screening for compounds having the ability to interfere

CC with the H. pylori life cycle or to inhibit H. pylori infection.

XX Sequence 579 BP; 173 A; 130 C; 98 G; 178 T; 0 other;

Query Match 0.6%; Score 19; DB 19; Length 579;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 409 acattattttacttgaa 427
 |||||||
 Db 356 ACAATATTTTTAACTTGA 338

RESULT 14
AA14445/c
ID AAX14445 standard; DNA: 649 BP.
XX
AC AAX14445;
XX
DT 31-MAR-1999 (first entry)
XX
DE H. pylori GHP0 875 gene.
XX
KM GHP0 protein; Helicobacter infection; gastroduodenal disease; gastritis;
KM peptic ulcer disease; ss.
XX
OS Helicobacter pylori.
XX
FH Key Location/Qualifiers
FT CDS 19..624
FT /tag= a
XX
PN MO9843478-A1.
XX
DR 08-OCT-1998.
XX
PP 01-APR-1998; 98MO-US06371.
XX
PR 29-JUL-1997; 97US-0902615.
PR 01-APR-1997; 97US-0833457.
PR 24-JUN-1997; 97US-0881227.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (IMMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX
PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
XX
DR WPI: 1998-542293/46.
DR P-PSDB; AAW98726.
XX
PT New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
PT infections and gastrointestinal diseases
XX
XX Claim 1; Page 1538; 2054pp; English.
XX
CC This sequence represents a polynucleotide of the invention. It was
CC isolated from Helicobacter pylori and encodes a H.pylori GHP0 protein.
CC The polypeptides can be used for preventing or treating Helicobacter
CC infections, and gastroduodenal diseases associated with these
CC infections, including acute, chronic, and atrophic gastritis, and peptic
CC ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used
CC for the production of antibodies. The products can also be used for
CC detection and diagnosis.
XX
SQ Sequence 649 BP; 191 A; 139 C; 118 G; 201 T; 0 other;

Query Match 0.6%; Score 19; DB 19; Length 649;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 409 acattatttacttgaa 427
|||||
DB 401 ACATTATTTTACTTGAA 383

RESULT 15
AA13098
ID AA13098 standard; cDNA: 1908 BP.
XX
AC AA13098;
XX
DT 13-MAR-2001 (first entry)
XX
DE Aspergillus oryzae EST SEQ ID NO:5621.

XX
KM Multiple gene expression; filamentous fungal cell; EST;
KM expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KM Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KM culture condition; environmental stress; spore morphogenesis;
KM metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS Aspergillus oryzae.
XX
PN WO200056762-A2.
XX
PD 28-SEP-2000.
XX
PF 22-MAR-2000; 2000MO-US07781.
XX
PR 22-MAR-1999; 99US-0273623.
XX
PA (NOVO) NOVO NORDISK BIOTECH INC.
PA (NOVO) NOVO NORDISK AS.
XX
PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX
DR WPI: 2000-594572/56.
XX
PT Monitoring differential expression of genes in filamentous fungal cells
PT uses fluorescence-labeled nucleic acids isolated from the cells and a
PT substrate of expressed sequence tags -
XX
PS Claim 88; Page 2332; 3161pp; English.

XX
CC The present invention describes a method for monitoring differential
CC expression of genes in a first filamentous fungal (FF) cell relative to
CC expression of the same genes in one or more second filamentous fungal
CC cells. The method uses fluorescence-labeled nucleic acids isolated from
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC are used in the methods for monitoring differential expression of genes
CC in a first filamentous fungal (FF) cell relative to expression of the
CC same genes in one or more second filamentous fungal cells. Monitoring
CC the global expression of genes from FF cells allows the production
CC potential of the microorganisms to be improved. New genes may be
CC discovered, possible functions of unknown open reading frames can be
CC identified and gene copy number variation and stability can be
CC monitored. The expression of genes can be used to study how FF cells
CC adapt to changes in culture conditions, environmental stress, spore
CC morphogenesis, recombination, metabolic or catabolic pathway
CC engineering. Using ESTs provides several advantages over genomic or
CC random cDNA clones including elimination of redundancy as one spot on an
CC array equals one gene or open reading frame, and organisation of the
CC microarrays based on function of the gene products to facilitate
CC analysis of the results. AA07478 to AA11247 represents ESTs from
CC Fusarium venenatum; AA11248 to AA11853 represents ESTs from
CC niger; AA11854 to AA14878 represents ESTs from Aspergillus oryzae; and
CC AA14879 to AA15337 represents ESTs from Trichoderma reesei, which are
CC all specifically claimed in the present invention.

XX
SQ Sequence 1908 BP; 441 A; 497 C; 466 G; 496 T; 8 other;

Query Match 0.6%; Score 19; DB 21; Length 1908;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 555 ggcagtcactctgtctgc 573
|||||
DB 883 ggcagtcactctgtctgc 901

RESULT 16
AA16202/c
ID AA16202 standard; cDNA: 2825 BP.
XX
AC AA16202;
XX


```

XX  MO9967421-A1.
XX  29-DEC-1999.
XX  25-JUN-1999; 99WO-N200092.
XX  25-JUN-1998; 98US-0105307.
XX  (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.
XX  (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX  Havukkala IJ, Bloksberg LN, Glenn M;
XX  WPI; 2000-116958/10.
XX  New plant microsatellite markers and associated flanking species for
XX  the detection of polymorphic genetic markers -
XX  Claim 1; Page 140-141; 392pp; English.
XX  Sequences AAA31040-A32093 represent novel plant microsatellite sequences
XX  and associated flanking species. The sequences comprise a central core
XX  repeat sequence, especially selected from the sequences AAA32094-A32096
XX  with left and right flanking sequences. The polynucleotide sequences
XX  can be used in the detection of DNA polymorphisms, in genome mapping,
XX  in physical mapping, in positional cloning of genes, in variety
XX  identification and in evaluation of genetic variability within and
XX  between plant tissues, populations, cultivars, species and species
XX  groups. They may also be used to design hybridization probes for
XX  oligonucleotide fingerprinting and library screening and to design
XX  primers for microsatellite-primed PCR. Microsatellite markers are
XX  useful to locate specific economically useful genes in plant genomes.
XX  Sequence 274 BP; 59 A; 79 C; 75 G; 60 T; 1 other:
XX  SQ
XX  Query Match 0.6%; Score 18; DB 21; Length 274;
XX  Best Local Similarity 100.0%; Pred. No. 1.3e+02;
XX  Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX  OY 3056 tgatgatgatgatctcag 3073
XX  |||||||
XX  DB 101 TGATGATGATGATCTCAG 84
XX  RESULT 19
XX  AAT24250/C
XX  ID AAT24250 standard; cDNA to mRNA; 322 BP.
XX  AAT24250;
XX  16-OCT-1996 (first entry)
XX  DE Human gene signature HUMGS06268.
XX  DE Gene signature; messenger RNA; mRNA; relative abundance; frequency;
XX  KW human; cloning; mapping; non-biased library; diagnosis; detection;
XX  KW cell typing; abnormal cell function; ss.
XX  OS Homo sapiens.
XX  PN WO9514772-A1.
XX  PD 01-JUN-1995.
XX  PF 11-NOV-1994; 94WO-JP01916.
XX  PR 12-NOV-1993; 93JP-0355504.
XX  PA (MATS/) MATSUBARA K.
XX  PA (OKUB/) OKUBO K.
XX  XX

```

```

PI  Matsubara K, Okubo K;
XX  DR WPI; 1995-206931/27.
XX  PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
XX  PT for diagnosis of abnormal cell function, by preparing cDNA that
XX  PT reflects relative abundance of corresp. mRNA in specific human
XX  PT tissues
XX  PS Claim 1; Page 1566; 2245pp; Japanese.
XX  CC A single-stranded DNA (or its complementary strand or the corresp.
XX  CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
XX  CC given in AAT19001-T26837 and which is able to hybridise to part of
XX  CC human genomic DNA, cDNA or mRNA is claimed. The GS (gene signature)
XX  CC sequences were obtained from 3'-directed cDNA libraries prepared
XX  CC from various human tissues; synthesis of cDNA was initiated from the
XX  CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
XX  CC untranslated sequence is unique to a particular mRNA species, almost
XX  CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
XX  CC is constructed so as to reflect accurately the relative abundance of
XX  CC different mRNAs in the particular tissue from which it was derived.
XX  CC The appearance frequency of a given GS in a cDNA library can be
XX  CC determined (esp. using primers and probes derived from the GS
XX  CC sequences) as a means of diagnosing abnormal cell function or for
XX  CC recognising different cell types.
XX  SQ Sequence 322 BP; 84 A; 53 C; 55 G; 119 T; 11 other:
XX  OY 36 atgaattcataaagac 53
XX  |||||||
XX  DB 51 ATGAATTCATTAAGGAC 34
XX  RESULT 20
XX  AAA31416/C
XX  ID AAA31416 standard; DNA; 355 BP.
XX  AAA31416;
XX  DT 05-JUL-2000 (first entry)
XX  DE Plant microsatellite marker #377.
XX  KW Plant microsatellite sequence; core repeat sequence; detection; probe;
XX  KW DNA polymorphism; genome mapping; physical mapping; fingerprinting;
XX  KW variety identification; genetic variability evaluation; primer; ss.
XX  OS Eucalyptus grandis.
XX  PN WO9967421-A1.
XX  PD 29-DEC-1999.
XX  PF 25-JUN-1999; 99WO-N200092.
XX  PR 25-JUN-1998; 98US-0105307.
XX  PA (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.
XX  PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX  PI Havukkala IJ, Bloksberg LN, Glenn M;
XX  DR WPI; 2000-116958/10.
XX  PT New plant microsatellite markers and associated flanking species for
XX  PT the detection of polymorphic genetic markers -
XX  XX

```


PS Claim 1; Page 188; 392pp; English.

CC Sequences AAA31040-A32093 represent novel plant microsatellite sequences and associated flanking species. The sequences comprise a central core repeat sequence, especially selected from the sequences AAA32094-A32096 with left and right flanking sequences. The polynucleotide sequences can be used in the detection of DNA polymorphisms, in genome mapping, in physical mapping, in positional cloning of genes, in variety identification and in evaluation of genetic variability within and between plant tissues, populations, cultivars, species and species groups. They may also be used to design hybridization probes for oligonucleotide fingerprinting and library screening and to design primers for microsatellite-primed PCR. Microsatellite markers are useful to locate specific economically useful genes in plant genomes.

XX

SQ Sequence 355 BP; 69 A; 98 C; 119 G; 68 T; 1 other;

Query Match 0.6%; Score 18; DB 21; Length 355;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3056 tgatgatgatgatctcag 3073
|||||
Db 101 TGATGATGATGATCTCAG 84

RESULT 21

AAA31325/C
ID AAA31325 standard; DNA; 377 BP.

XX

AC AAA31325;

XX

DT 05-JUL-2000 (first entry)

XX

DE Plant microsatellite marker #286.

XX

KW Plant microsatellite sequence; core repeat sequence; detection; probe; DNA polymorphism; genome mapping; physical mapping; fingerprinting; variety identification; genetic variability evaluation; primer; ss.

XX

OS Eucalyptus grandis.

XX

PN WO9967421-A1.

XX

PD 29-DEC-1999.

XX

PT 25-JUN-1999; 99WO-NZ00092.

XX

PS 25-JUN-1998; 98US-0105307.

XX

PA (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.
(FLET-) FLETCHER CHALLENGE FORESTS LTD.

XX

PI Havukkala JU, Bloksberg LN, Glenn M;

XX

DR WPI; 2000-116958/10.

XX

PT New plant microsatellite markers and associated flanking species for the detection of polymorphic genetic markers -

XX

PS Claim 1; Page 161; 392pp; English.

CC Sequences AAA31040-A32093 represent novel plant microsatellite sequences and associated flanking species. The sequences comprise a central core repeat sequence, especially selected from the sequences AAA32094-A32096 with left and right flanking sequences. The polynucleotide sequences can be used in the detection of DNA polymorphisms, in genome mapping, in physical mapping, in positional cloning of genes, in variety identification and in evaluation of genetic variability within and between plant tissues, populations, cultivars, species and species groups. They may also be used to design hybridization probes for oligonucleotide fingerprinting and library screening and to design

CC primers for microsatellite-primed PCR. Microsatellite markers are useful to locate specific economically useful genes in plant genomes.

XX

SQ Sequence 377 BP; 72 A; 102 C; 130 G; 73 T; 0 other;

Query Match 0.6%; Score 18; DB 21; Length 377;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3056 tgatgatgatgatctcag 3073
|||||
Db 88 TGATGATGATGATCTCAG 71

RESULT 22

AAA31366/C
ID AAA31366 standard; DNA; 435 BP.

XX

AC AAA31366;

XX

DT 05-JUL-2000 (first entry)

XX

DE Plant microsatellite marker #327.

XX

KW Plant microsatellite sequence; core repeat sequence; detection; probe; DNA polymorphism; genome mapping; physical mapping; fingerprinting; variety identification; genetic variability evaluation; primer; ss.

XX

OS Eucalyptus grandis.

XX

PN WO9967421-A1.

XX

PD 29-DEC-1999.

XX

PT 25-JUN-1999; 99WO-NZ00092.

XX

PS 25-JUN-1998; 98US-0105307.

XX

PA (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.
(FLET-) FLETCHER CHALLENGE FORESTS LTD.

XX

PI Havukkala JU, Bloksberg LN, Glenn M;

XX

DR WPI; 2000-116958/10.

XX

PT New plant microsatellite markers and associated flanking species for the detection of polymorphic genetic markers -

XX

PS Claim 1; Page 173; 392pp; English.

CC Sequences AAA31040-A32093 represent novel plant microsatellite sequences and associated flanking species. The sequences comprise a central core repeat sequence, especially selected from the sequences AAA32094-A32096 with left and right flanking sequences. The polynucleotide sequences can be used in the detection of DNA polymorphisms, in genome mapping, in physical mapping, in positional cloning of genes, in variety identification and in evaluation of genetic variability within and between plant tissues, populations, cultivars, species and species groups. They may also be used to design hybridization probes for oligonucleotide fingerprinting and library screening and to design primers for microsatellite-primed PCR. Microsatellite markers are useful to locate specific economically useful genes in plant genomes.

XX

SQ Sequence 435 BP; 85 A; 125 C; 146 G; 79 T; 0 other;

Query Match 0.6%; Score 18; DB 21; Length 435;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3056 tgatgatgatgatctcag 3073
|||||

Db 101 TGATGATGATCTCAG 84

RESULT 23
AAC09333/C
ID AAC09333 standard; cDNA: 446 BP.
XX
XX AAC09333;
XX
XX 06-OCT-2000 (first entry)
XX
XX Human secreted protein 5' EST, SEQ ID NO: 13408.
DE
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
XX Homo sapiens.
OS
XX EP1033401-A2.
PN
XX 06-SEP-2000.
PD
XX 21-FEB-2000; 2000EP-0200610.
XX
XX 26-FEB-1999; 9905-0122487.
XX
XX (GSEST) GENSET.
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J;
PI WPI: 2000-500381/45.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX Claim 1; SEQ ID 13408; 71pp + CD-ROM; English.
PS
XX The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.
XX
XX Sequence 446 BP; 134 A; 76 C; 70 G; 162 T; 4 other;

Query Match 0.6%; Score 18; DB 21; Length 446;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 36 atgaattcataagac 53
|||||
Db 106 ATGAATTCATTAAGAC 89

RESULT 24
AA114905/C
ID AA114905 standard; DNA: 454 BP.
XX
XX AA114905;
XX
XX 12-OCT-2001 (first entry)
XX

DE Probe #4838 for gene expression analysis in human cervical cell sample.
XX
XX Probe: human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.
XX
XX Homo sapiens.
OS
XX WO200157278-A2.
PN
XX 09-AUG-2001.
PD
XX 30-JAN-2001; 2001WO-US00670.
PF
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI WPI: 2001-488901/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
PT
XX
XX Claim 25; SEQ ID No 4838; 487pp; English.
PS
XX The present invention relates to human single exon nucleic acid probes
CC (SENPs). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 454 BP; 119 A; 89 C; 110 G; 136 T; 0 other;

Query Match 0.6%; Score 18; DB 22; Length 454;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 540 ggggaatctggcaagac 557
|||||
Db 29 GGGGAATCTGGCAAGAC 12

RESULT 25
AA136257/C
ID AA136257 standard; DNA: 454 BP.
XX
XX AA136257;
AC
XX 17-OCT-2001 (first entry)
DT
XX Probe #4943 used to measure gene expression in human placenta sample.
DE
XX Probe: microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
XX Homo sapiens.
OS
XX WO200157272-A2.
PN
XX 09-AUG-2001.
PD

```
XX 30-JAN-2001; 2001WO-US00663.
PF
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID No 4943; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 454 BP; 119 A; 89 C; 110 G; 136 T; 0 other;

Query Match          0.6%; Score 18; DB 22; Length 454;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 540 ggggaatctggcaagc 557
   |||||||
DB 29 GGGGAATCTGCAAGGC 12

RESULT 26
AA104685/C
ID AA104685 standard; DNA; 454 BP.
XX
AC AA104685;
XX
TT 09-OCT-2001 (first entry)
XX
DE Probe #4676 used to measure gene expression in human breast sample.
XX
KM Probe: human; breast disease; breast cancer; development disorder; ss;
XX inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
OS Homo sapiens.
XX
PN WO200157270-A2.
XX
PD 09-AUG-2001.
XX
PI 29-JAN-2001; 2001WO-US00661.
XX
PF
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
```

```
DR WPI; 2001-476286/51.
XX
XX Novel single exon nucleic acid probe used to measuring gene expression
PT in a human breast -
XX
PS Claim 25; SEQ ID No 4676; 322pp; English.
XX
CC The present invention relates to novel single exon nucleic acid probes.
CC The present sequence is one such probe. The probes are useful for
CC measuring human gene expression in a human breast sample, where the probe
CC hybridises at high stringency to a nucleic acid expressed in the human
CC breast. The probes are useful for predicting, diagnosing, grading,
CC staging, monitoring and prognosing diseases of the human breast,
CC particularly those diseases with polygenic aetiology. The diseases
CC include: breast cancer, disorders of development, inflammatory diseases
CC of the breast, fibrocystic changes, proliferative breast disease and
CC non-carcinoma tumours.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 454 BP; 119 A; 89 C; 110 G; 136 T; 0 other;

Query Match          0.6%; Score 18; DB 22; Length 454;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 540 ggggaatctggcaagc 557
   |||||||
DB 29 GGGGAATCTGCAAGGC 12

RESULT 27
AAC01809
ID AAC01809 standard; CDNA; 466 BP.
XX
AC AAC01809;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 1807.
XX
KM Human: 5' EST; expressed sequence tag; secreted protein; CDNA isolation;
XX gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST ) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.
DR P-PSDB; AAC01803.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 1; SEQ ID 1807; 71pp + CD-ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
```

CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors.
XX
SQ Sequence 466 BP; 76 A; 106 C; 132 G; 148 T; 4 other;

Query Match 0.6%; Score 18; DB 21; Length 466;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 766 ttctcctctgtagcgt 763
|||||
DB 220 ttctcctctgtagcgt 237

RESULT 28
ID AAA31287/C
AAA31287 standard; DNA; 468 BP.
XX
AC AAA31287;
XX
DT 05-JUN-2000 (first entry)
XX
DE Plant microsatellite marker #248.
XX
KW Plant microsatellite sequence; core repeat sequence; detection; probe;
KW DNA polymorphism; genome mapping; physical mapping; fingerprinting;
KW variety identification; genetic variability evaluation; primer; ss.
XX
OS Eucalyptus grandis.
XX
PN WO9967421-A1.
XX
PD 29-DEC-1999.
XX
PF 25-JUN-1999; 99WO-NZ00092.
XX
PR 25-JUN-1998; 98US-0105307.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.
XX (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PA Havukkala II, Bloksberg LN, Glenn M;
XX
DR WPI; 2000-116958/10.
XX
PT New plant microsatellite markers and associated flanking species for
PT the detection of polymorphic genetic markers -
XX
PS Claim 1; Page 149; 392pp; English.
XX
CC Sequences AAA31040-A32093 represent novel plant microsatellite sequences
CC and associated flanking species. The sequences comprise a central core
CC repeat sequence, especially selected from the sequences AAA32094-A32096
CC with left and right flanking sequences. The polynucleotide sequences
CC can be used in the detection of DNA polymorphisms, in genome mapping,
CC in physical mapping, in positional cloning of genes, in variety
CC identification and in evaluation of genetic variability within and
CC between plant tissues, populations, cultivars, species and species
CC groups. They may also be used to design hybridization probes for
CC oligonucleotide fingerprinting and library screening and to design
CC primers for microsatellite-primed PCR. Microsatellite markers are
CC useful to locate specific economically useful genes in plant genomes.
XX
SQ Sequence 468 BP; 95 A; 129 C; 156 G; 87 T; 1 other;

Query Match 0.6%; Score 18; DB 21; Length 468;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3056 tgatgatgatgctcag 3073
|||||
DB 101 TGATGATGATGATCAG 84

RESULT 29
ID AAI1787
AAI1787 standard; DNA; 472 BP.
XX
AC AAI1787;
XX
DT 12-OCT-2001 (first entry)
XX
DE Probe #1720 for gene expression analysis in human cervical cell sample.
XX
KW Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157278-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00670.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX
PS Claim 25; SEQ ID No 1720; 487pp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENPs). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 472 BP; 129 A; 77 C; 153 G; 113 T; 0 other;

Query Match 0.6%; Score 18; DB 22; Length 472;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 166 agaagtgagcagaagc 183
|||||
DB 143 agaagtgagcagaagc 160

```

RESULT 30
AAI33101
ID AAI33101 standard; DNA; 472 BP.
XX
AC AAI33101;
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #1787 used to measure gene expression in human placenta sample.
XX
KW Probe: microarray; human; placenta; antenatal diagnosis;
XX genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-468897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID No 1787; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP).
XX The present sequence is one such probe. The probes are useful for
XX producing a microarray for predicting, measuring and displaying gene
XX expression in samples derived from human placenta. The probes are useful
XX for antenatal diagnosis of human genetic disorders.
XX
Q Sequence 472 BP; 129 A; 77 C; 153 G; 113 T; 0 other;
XX
Query Match 0.6%; Score 18; DB 22; Length 472;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 166 agaaggtgagcagatg 183
Db 143 agaaggtgagcagatg 160
XX
RESULT 31
AAI01718
ID AAI01718 standard; DNA; 472 BP.
XX
AC AAI01718;
XX
DT 09-OCT-2001 (first entry)
XX
DE Probe #1709 used to measure gene expression in human breast sample.
XX
KW Probe: human; breast disease; breast cancer; development disorder; ss;
XX inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
OS Homo sapiens.

```

```

XX
XX WO200157270-A2.
XX
PD 09-AUG-2001.
XX
PF 29-JAN-2001; 2001WO-US00661.
XX
PR 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-476286/51.
XX
PT Novel single exon nucleic acid probe used to measuring gene expression
XX in a human breast -
XX
PS Claim 25; SEQ ID No 1709; 322pp; English.
XX
CC The present invention relates to novel single exon nucleic acid probes.
XX The present sequence is one such probe. The probes are useful for
XX measuring human gene expression in a human breast sample, where the probe
XX hybridises at high stringency to a nucleic acid expressed in the human
XX breast. The probes are useful for predicting, diagnosing, grading,
XX staging, monitoring and prognosing diseases of the human breast,
XX particularly those diseases with polygenic aetiology. The diseases
XX include: breast cancer, disorders of development, inflammatory diseases
XX of the breast, fibrocystic changes, proliferative breast disease and
XX non-carcinoma tumours.
XX
CC Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 472 BP; 129 A; 77 C; 153 G; 113 T; 0 other;
XX
Query Match 0.6%; Score 18; DB 22; Length 472;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 166 agaaggtgagcagatg 183
Db 143 agaaggtgagcagatg 160
XX
RESULT 32
AAQ05868/C
ID AAQ05868 standard; DNA; 594 BP.
XX
AC AAQ05868;
XX
DT 07-JAN-1991 (first entry)
XX
DE Sequence encoding mamalian growth hormone receptor binding protein.
XX
KW Ovine placental lactogen; ds;
XX
XX Ovis ammon aries.
XX
PN EP386979-A.
XX
PD 12-SEP-1990.
XX
PF 05-MAR-1990; 90EP-0302322.
XX
PR 06-MAR-1989; 89US-0319585.

```


PR 04-MAY-1999; 9905-0132484.
PR 05-MAY-1999; 9905-0132485.
PR 06-MAY-1999; 9905-0132486.
PR 07-MAY-1999; 9905-0132487.
PR 11-MAY-1999; 9905-0132483.
PR 14-MAY-1999; 9905-0134218.
PR 14-MAY-1999; 9905-0134219.
PR 14-MAY-1999; 9905-0134221.
PR 14-MAY-1999; 9905-0134370.
PR 18-MAY-1999; 9905-0134768.
PR 19-MAY-1999; 9905-0134941.
PR 20-MAY-1999; 9905-0135124.
PR 21-MAY-1999; 9905-0135353.
PR 24-MAY-1999; 9905-0135529.
PR 25-MAY-1999; 9905-0136021.
PR 27-MAY-1999; 9905-0136392.
PR 28-MAY-1999; 9905-0136782.
PR 01-JUN-1999; 9905-0137222.
PR 03-JUN-1999; 9905-0137528.
PR 04-JUN-1999; 9905-0137502.
PR 07-JUN-1999; 9905-0137724.
PR 08-JUN-1999; 9905-0138094.
PR 10-JUN-1999; 9905-0138540.
PR 14-JUN-1999; 9905-0138847.
PR 16-JUN-1999; 9905-0139452.
PR 16-JUN-1999; 9905-0139453.
PR 17-JUN-1999; 9905-0139492.
PR 18-JUN-1999; 9905-0139454.
PR 18-JUN-1999; 9905-0139455.
PR 18-JUN-1999; 9905-0139456.
PR 18-JUN-1999; 9905-0139457.
PR 18-JUN-1999; 9905-0139458.
PR 18-JUN-1999; 9905-0139459.
PR 18-JUN-1999; 9905-0139460.
PR 18-JUN-1999; 9905-0139461.
PR 18-JUN-1999; 9905-0139462.
PR 18-JUN-1999; 9905-0139463.
PR 18-JUN-1999; 9905-0139750.
PR 18-JUN-1999; 9905-0139763.
PR 21-JUN-1999; 9905-0139817.
PR 22-JUN-1999; 9905-0139899.
PR 23-JUN-1999; 9905-0140353.
PR 23-JUN-1999; 9905-0140354.
PR 24-JUN-1999; 9905-0140695.
PR 28-JUN-1999; 9905-0140823.
PR 29-JUN-1999; 9905-0140991.
PR 30-JUN-1999; 9905-0141287.
PR 01-JUL-1999; 9905-0141842.
PR 01-JUL-1999; 9905-0142154.
PR 02-JUL-1999; 9905-0142055.
PR 06-JUL-1999; 9905-0142390.
PR 08-JUL-1999; 9905-0142803.
PR 09-JUL-1999; 9905-0142977.
PR 12-JUL-1999; 9905-0142920.
PR 13-JUL-1999; 9905-0143542.
PR 14-JUL-1999; 9905-0143624.
PR 15-JUL-1999; 9905-0144005.
PR 16-JUL-1999; 9905-0144085.
PR 16-JUL-1999; 9905-0144086.
PR 19-JUL-1999; 9905-0144325.
PR 19-JUL-1999; 9905-0144331.
PR 19-JUL-1999; 9905-0144332.
PR 19-JUL-1999; 9905-0144333.
PR 19-JUL-1999; 9905-0144334.
PR 19-JUL-1999; 9905-0144335.
PR 20-JUL-1999; 9905-0144352.
PR 20-JUL-1999; 9905-0144632.
PR 21-JUL-1999; 9905-0144884.
PR 21-JUL-1999; 9905-0145086.
PR 21-JUL-1999; 9905-0145088.

PR 22-JUL-1999; 9905-0145085.
PR 22-JUL-1999; 9905-0145087.
PR 22-JUL-1999; 9905-0145089.
PR 22-JUL-1999; 9905-0145192.
PR 23-JUL-1999; 9905-0145145.
PR 23-JUL-1999; 9905-0145218.
PR 23-JUL-1999; 9905-0145224.
PR 26-JUL-1999; 9905-0145276.
PR 27-JUL-1999; 9905-0145913.
PR 27-JUL-1999; 9905-0145918.
PR 27-JUL-1999; 9905-0145919.
PR 28-JUL-1999; 9905-0145951.
PR 02-AUG-1999; 9905-0146386.
PR 02-AUG-1999; 9905-0146388.
PR 03-AUG-1999; 9905-0146389.
PR 04-AUG-1999; 9905-0147038.
PR 04-AUG-1999; 9905-0147204.
PR 05-AUG-1999; 9905-0147302.
PR 05-AUG-1999; 9905-0147192.
PR 06-AUG-1999; 9905-0147260.
PR 06-AUG-1999; 9905-0147303.
PR 06-AUG-1999; 9905-0147416.
PR 09-AUG-1999; 9905-0147493.
PR 09-AUG-1999; 9905-0147935.
PR 10-AUG-1999; 9905-0148171.
PR 11-AUG-1999; 9905-0148319.
PR 12-AUG-1999; 9905-0148341.
PR 13-AUG-1999; 9905-0148565.
PR 13-AUG-1999; 9905-0148684.
PR 16-AUG-1999; 9905-0149368.
PR 17-AUG-1999; 9905-0149175.
PR 18-AUG-1999; 9905-0149426.
PR 20-AUG-1999; 9905-0149722.
PR 20-AUG-1999; 9905-0149723.
PR 20-AUG-1999; 9905-0149929.
PR 23-AUG-1999; 9905-0149902.
PR 23-AUG-1999; 9905-0149930.
PR 25-AUG-1999; 9905-0150566.
PR 26-AUG-1999; 9905-0150884.
PR 27-AUG-1999; 9905-0151065.
PR 27-AUG-1999; 9905-0151066.
PR 27-AUG-1999; 9905-0151080.
PR 30-AUG-1999; 9905-0151303.
PR 31-AUG-1999; 9905-0151438.
PR 01-SEP-1999; 9905-0151930.
PR 07-SEP-1999; 9905-0152363.
PR 10-SEP-1999; 9905-0153070.
PR 13-SEP-1999; 9905-0153758.
PR 16-SEP-1999; 9905-0154018.
PR 16-SEP-1999; 9905-0154039.
PR 20-SEP-1999; 9905-0154779.
PR 22-SEP-1999; 9905-0155139.
PR 23-SEP-1999; 9905-0155486.
PR 24-SEP-1999; 9905-0155659.
PR 24-SEP-1999; 9905-0156458.
PR 29-SEP-1999; 9905-0156596.
PR 04-OCT-1999; 9905-0157117.
PR 05-OCT-1999; 9905-0157753.
PR 06-OCT-1999; 9905-0157865.
PR 07-OCT-1999; 9905-0158029.
PR 08-OCT-1999; 9905-0158232.
PR 12-OCT-1999; 9905-0158369.
PR 13-OCT-1999; 9905-0159293.
PR 13-OCT-1999; 9905-0159294.
PR 13-OCT-1999; 9905-0159295.
PR 13-OCT-1999; 9905-0159295.
PR 14-OCT-1999; 9905-0159330.
PR 14-OCT-1999; 9905-0159331.
PR 14-OCT-1999; 9905-0159637.
PR 14-OCT-1999; 9905-0159638.
PR 18-OCT-1999; 9905-0159584.
PR 21-OCT-1999; 9905-0160741.
PR 21-OCT-1999; 9905-0160767.

CC sequences can be used for detecting Enterococcus nucleic acids.
 CC Products from the present invention can also be used for screening
 CC compounds to identify agonists and antagonists of E. faecalis protein
 CC activity.

Sequence 888 BP; 290 A; 159 C; 186 G; 253 T; 0 other;

Query Match 0.6%; Score 18; DB 20; Length 888;

Best Local Similarity 100.0%; Pred. No. 1.4e+02; Mismatches 0; Indels 0; Gaps 0;

QY 2527 gctgtctgcaatgcag 2544
 ||||||||||||||||
 DB 110 gctgtctgcaatgcag 127

RESULT 37

AAH67086
 ID AAH67086 standard; DNA; 960 BP.

XX AAH67086;

DT 26-SEP-2001 (first entry)

DE C glutamicum coding sequence fragment SEQ ID NO: 2121.

KM Corynebacterium; amino acid synthesis; vitamin; saccharide;
 KW organic acid synthesis; ds.

XX Corynebacterium glutamicum.

PN EP1108790-A2.

PD 20-JUN-2001.

PF 18-DEC-2000; 2000EP-0127688.

PR 16-DEC-1999; 99JP-0377484.

PR 07-APR-2000; 2000JP-0159162.

PR 03-AUG-2000; 2000JP-0280988.

XX (KYOWA) KYOWA HAKKO KOGYO KK.

XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

RI Tateishi N, Senoh A, Ikeda M, Ozaki A;

DR WPI; 2001-376931/40.

XX P-PSDB; AAG91867.

Claim 8: SEQ ID NO: 2121; 246pp + Sequence listing; English.

CC The present invention provides a number of nucleotide and protein
 CC sequences from the Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of Corynebacterium glutamicum, measuring expression amount and
 CC analyzing the expression profile or expression pattern of a gene derived
 CC from Corynebacterium glutamicum, and identifying a homologue of a gene
 CC from Corynebacterium glutamicum. Corynebacterium bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a nucleic acid described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.

Sequence 960 BP; 206 A; 235 C; 271 G; 248 T; 0 other;

Query Match 0.6%; Score 18; DB 22; Length 960;
 CC Best Local Similarity 100.0%; Pred. No. 1.4e+02; Mismatches 0; Indels 0; Gaps 0;

QY 3040 ttgttggtgcaattg 3057
 ||||||||||||||||
 DB 394 ttgttggtgcaattg 411

RESULT 38

AAQ05870/C
 ID AAQ05870 standard; cDNA; 992 BP.

XX AAQ05870;

DT 07-JAN-1991 (first entry)

DE Sequence encoding mammalian growth hormone receptor binding protein.

KW Ovine placental lactogen; ds;

XX Ovis ammon aries.

XX Key Location/Qualifiers

FT CDS 69..776

XX BP386979-A.

PD 12-SEP-1990.

PF 05-MAR-1990; 90EP-0302322.

PR 06-MAR-1989; 89US-0319585.

XX (GENTH) GENENTECH INC.

XX Wood WI, Colosi PC;

DR WPI; 1990-276995/37.

DR P-PSDB; AAR06643.

XX DNA sequence encoding growth-hormone-receptor-binding protein -

PT useful for increase of mammalian meat and milk prodn. and

PT decrease in fat content.

XX Claim 14; Fig 3; 22pp; English.

CC Receptor binding protein may be used to increase meat and milk

CC production and decrease fat content especially in sheep and goats.

CC Sequence shows an identity of 26% with human growth hormone and 67%

CC to ovine placental lactogen.

CC Sequence was obtained using probe described in AAQ05869.

Sequence 992 BP; 308 A; 237 C; 195 G; 252 T; 0 other;

Query Match 0.6%; Score 18; DB 11; Length 992;
 CC Best Local Similarity 100.0%; Pred. No. 1.4e+02; Mismatches 0; Indels 0; Gaps 0;

QY 535 ttgaagggaatctgca 552
 ||||||||||||||||
 DB 236 ttgaagggaatctgca 219

RESULT 39

AAZ52527/C
 ID AAZ52527 standard; cDNA; 1034 BP.

XX AAZ52527;

XX

29-FEB-2000 (first entry)

Human secreted protein clone ya66_1 nucleotide sequence SEQ ID NO:105.

Human; secreted protein; immunostimulatory; haemostatic; cytokine; proliferative; differentiative; chemotactic; chemokinetic; vaccine; thrombolytic; antiinflammatory; cytostatic; immunosuppressive; gene therapy; ss.

Homo sapiens.

WO958642-A2.

18-NOV-1999.

14-MAY-1999; 99WO-US10843.

14-MAY-1998; 98US-0085472.

17-AUG-1998; 98US-0096824.

11-SEP-1998; 98US-0099843.

11-SEP-1998; 98US-0099950.

15-SEP-1998; 98US-0100424.

29-SEP-1998; 98US-0102329.

09-OCT-1998; 98US-0103615.

11-DEC-1998; 98US-0111799.

14-DEC-1998; 98US-0112159.

31-DEC-1998; 98US-0114415.

10-FEB-1999; 99US-0248059.

06-APR-1999; 99US-0287150.

13-MAY-1999; 99US-0311021.

(GEMV) GENETICS INST INC.

Mong GG, Clark HF, Fechtel K, Agostino MJ; WPI: 2000-053095/04. P-PSDB; AAY73442.

Novel polynucleotides and proteins having biological activities which make suitable for treating, preventing or ameliorating medical conditions in humans or animals -

Claim 114; Page 653; 730pp; English.

The present invention describes human secreted proteins encoded by polynucleotides obtained from adult testes, foetal brain, adult brain, brain (foetal and adult), foetal kidney, adult spleen, and adult thymus cDNA libraries. The polynucleotides and proteins are predicted to have biological activities which would make them suitable for treating, preventing or ameliorating medical conditions in humans and animals. Suggested activities include nutritional activity, cytokine and cell proliferation/differentiation activity, immune stimulating (e.g. as vaccines) or suppressing activity, haematopoiesis regulating activity, tissue growth activity, activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, receptor/ligand activity, anti-inflammatory activity, cadherin/tumour invasion suppressor activity, and tumour inhibition activity. The polynucleotides are also stated to be useful for gene therapy. Therapeutic compositions are also presently valuable for veterinary applications. AA252475 to CC AA252581 encode human secreted proteins, and AAY73390 to AAY73500 CC represent human secreted proteins, given in the present invention.

Sequence 1034 BP; 305 A; 228 C; 222 G; 276 T; 3 other;

Query Match 0.6%; Score 18; DB 21; Length 1034; Best Local Similarity 100.0%; Pred. No. 1.4e+02; Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

450 ctgtcgaagaagaccac 467
|||||
168 CTGTGAGAGAGACCA 151

RESULT 40
AAE71387
ID AAE71387 standard; DNA; 1083 BP.
XX
AC AAE71387;
XX
DT 30-APR-2001 (first entry)
XX
DE Corynebacterium glutamicum SMP protein nucleotide sequence SEQ ID NO:55.
XX
KW Corynebacterium glutamicum; carbon metabolism and energy production; SMP protein; sugar metabolism and oxidative phosphorylation protein; fine chemical production; organic acid; proteinogenic amino acid; nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside; nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol; carbohydrate; aromatic compound; vitamin; cofactor; polyketide; enzyme; diagnosis; Corynebacterium diphtheriae; evolutionary study; ds.
XX
OS Corynebacterium glutamicum.
XX
PN WO200100844-A2.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-IB00943.
XX
PR 25-JUN-1999; 99US-0141031.
PR 08-JUL-1999; 99DE-1031412.
PR 08-JUL-1999; 99DE-1031413.
PR 08-JUL-1999; 99DE-1031419.
PR 08-JUL-1999; 99DE-1031420.
PR 08-JUL-1999; 99DE-1031424.
PR 08-JUL-1999; 99DE-1031428.
PR 08-JUL-1999; 99DE-1031431.
PR 08-JUL-1999; 99DE-1031433.
PR 08-JUL-1999; 99DE-1031434.
PR 08-JUL-1999; 99DE-1031510.
PR 08-JUL-1999; 99DE-1031562.
PR 08-JUL-1999; 99DE-1031634.
PR 09-JUL-1999; 99DE-1032180.
PR 09-JUL-1999; 99DE-1032227.
PR 09-JUL-1999; 99DE-1032230.
PR 09-JUL-1999; 99US-0143208.
PR 14-JUL-1999; 99DE-1032924.
PR 14-JUL-1999; 99DE-1032973.
PR 14-JUL-1999; 99DE-1033005.
PR 27-AUG-1999; 99DE-1040765.
PR 31-AUG-1999; 99US-0151572.
PR 03-SEP-1999; 99DE-1042076.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042086.
PR 03-SEP-1999; 99DE-1042087.
PR 03-SEP-1999; 99DE-1042088.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042123.
PR 03-SEP-1999; 99DE-1042125.
XX
PA (BADI) BASF AG.
XX
PI Pompejus M, Kroegeer B, Schroeder H, Zelder O, Habernauer G;
XX
DR WPI: 2001-061975/07.
XX
DR P-PSDB; AAB79270.
XX
PT New isolated Corynebacterium glutamicum nucleic acid encoding a sugar
PT metabolism and oxidative phosphorylation protein for production or
PT modulation of production of fine chemicals e.g. amino acids,
PT carbohydrates or enzymes -
XX
PS Claim 3; Page 217-219; 1246pp; English.
XX
CC AAF71360 to AAE71750 encode the Corynebacterium glutamicum sugar

CC metabolism and oxidative phosphorylation (SMP) proteins given in
 CC AAB9243 to AAB 79633 which are involved in carbon metabolism and
 CC energy production. The C. glutamicum SMP gene can be used in vectors
 CC (II) for expression in host cells and production or modulation of
 CC production of fine chemicals, such as, an organic acid, a proteinogenic
 CC or nonproteinogenic amino acid (preferred), a purine or pyrimidine base,
 CC a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty
 CC acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a
 CC cofactor, a polyketide, or an enzyme. The presence of (I) or SMP proteins
 CC (III) encoded by them are used for diagnosing the presence or activity of
 CC Corynebacterium diphtheriae in a subject. (I), (II), (III) or host cells
 CC containing them are used to map genomes of organisms related to
 CC C. glutamicum, identify and localise C. glutamicum sequences of interest,
 CC in evolutionary studies, in determining SMP protein regions required
 CC for function, in modulating SMP protein activity, in modulating the
 CC metabolism of sugars, and in modulating high-energy molecule production
 CC in a cell (i.e. ATP, NADPH).

Sequence 1083 BP; 241 A; 261 C; 308 G; 273 T; 0 other;

Query Match

Best Local Similarity 100.0%; Score 18; DB 22; Length 1083;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3040 ttgttggtggtgcaattg 3057

DB 494 ttgttggtggtgcaattg 511

RESULT 41

AAZ42236
 ID AAZ42236 standard; cDNA; 1205 BP.

XX AAZ42236;

DT 31-JAN-2000 (first entry)

XX Human normal bladder tissue cDNA derived EST 115.

KW Human; bladder; treatment; EST; expressed sequence tag; cytostatic;

XX cancer; gene therapy; ss.

XX Homo sapiens.

PN DE19818620-A1.

XX 28-OCT-1999.

PF 21-APR-1998; 98DE-1018620.

PR 21-APR-1998; 98DE-1018620.

PA (META-) METAGEN CBS GENOMFORSCHUNG MBH.

PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;

DR MPI; 1999-602416/52.

DR P-PsDB; AAY60559, AAY60560, AAY60561.

PT New polypeptides and their nucleic acids, useful for treatment of
 bladder tumour and identification of therapeutic agents -

PS Claim 3; Page 235; 366pp; German.

XX This invention describes novel polypeptide fragment sequences (I) and
 CC their encoding nucleic acids (II) which are highly expressed in normal
 CC bladder tissue and have cytostatic activity. (II) are used for
 CC recombinant expression of (I) and to isolate complete genes. (I) are
 CC used to identify agents suitable for the treatment of bladder tumours,
 CC to directly treat this form of cancer (including expression from gene
 CC therapy vectors), or are used in a preparation for cancer treatment. (I)
 CC is also used for the generation of specific antibodies. (II) are

CC identified by assembling ESTs (expressed sequence tags) from a
 CC particular tissue type before comparison of expression patterns. This
 CC allows a significantly longer fragment of the gene to be revealed, and
 CC therefore reduces the number of failures because of ESTs from different
 CC libraries representing different parts of the same unknown gene
 CC distorting the estimated frequency of occurrence in a particular tissue.
 CC AAZ42122-242248 represent EST fragments derived from a human normal
 CC bladder tissue cDNA library which encode the protein fragments
 CC represented in AAY60329-Y60591.

Sequence 1205 BP; 274 A; 349 C; 338 G; 244 T; 0 other;

Query Match

Best Local Similarity 100.0%; Score 18; DB 20; Length 1205;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2016 actctgaggtcacacac 2033

DB 714 actctgaggtcacacac 731

RESULT 42

AAAN80316
 ID AAAN80316 standard; DNA; 1250 BP.

XX AAAN80316;

DT 10-JAN-1991 (first entry)

XX Transcription control region of gene encoding serine protease
 DE (SP) of human myeloid cell origin.

XX Serine protease; myeloid cell; intravascular coagulation treatment; ss.

XX Homo sapiens.

PN WO8806621-A.

XX 07-SEP-1988.

PF 26-FEB-1988; 88WO-JP00205.

PR 09-SEP-1987; 87JP-0225540.

PR 05-MAR-1987; 87JP-0050676.

XX (TORA) TORAY IND INC (AOKI/).

PI Aoki Y, Okano K, Naruto M, Shimizu H, Nakamura H;

DR WPI; 1988-271161/38.

PT Serine protease of human myeloid cell origin -
 useful in disseminated intravascular coagulation treatment

PS Disclosure; Fig 15; 52pp; Japanese.

XX SP DNA is selected and cloned from a human myeloid cell library,
 CC inserted in a suitable vector and expressed in a transformant of a
 CC suitable organism such as E. coli HMS-174 or HB-101 or a yeast. It is new
 CC so are a SP of human myeloid cell origin; precursors of SP having an
 CC N-terminal splittable or signal peptide; and a transcription-controlling
 CC DNA sequence required for expression of the gene. SP has antithrombotic
 CC activity, esp. useful in disseminated intravascular coagulation. By
 CC expression of the protease gene in a suitable transformant organism SP
 CC can be obtained in large quantity.

Sequence 1250 BP; 305 A; 340 C; 385 G; 220 T; 0 other;

Query Match

Best Local Similarity 100.0%; Score 18; DB 9; Length 1250;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1081 tgggtgaagtgaattcc 1098
 |||||
Db 43 tgggtgaagtgaattcc 60

RESULT 43
AAC40012
ID AAC40012 standard; DNA: 1473 BP.
XX AAC40012;
AC
XX 17-OCT-2000 (first entry)
DT
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 26727.
DE
XX
XX Hybridisation assay: genetic mapping: gene expression control;
KM protein identification; signal transduction pathway;
KM metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132485.
PR 05-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136332.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.


```

XX OS Ctenocephalides felis.
XX XX
XX PN WO9821324-A1.
XX XX
XX PD 22-MAY-1998.
XX XX
XX PF 10-NOV-1997; 97WO-US20598.
XX XX
XX PR 12-NOV-1996; 96US-0747221.
XX XX
XX PA (HESK-) HESKA CORP.
XX XX
XX PI Brandt KS, Silver GM, Wisniewski N;
XX XX
XX DR WPI; 1998-297929/26.
XX DR P-PSDB; AAW57852.
XX XX
XX PT New nucleic acid encoding carboxyl:esterase(s) from fleas - useful,
XX PT e.g. in vaccines, for preventing infestation by haematophagous
XX PT ectoparasites, particularly on cats and dogs
XX XX
XX (S) Claim 1; Page 113-115; 230pp; English.
XX XX
XX CC This sequence encodes the flea esterase protein, nfe51515 (the
XX CC complementary strand is shown in AAV40738), of the invention. When
XX CC administered to animals, the protein induces a protective immune
XX CC (antibody) response against carboxylesterase (CE), so they, or
XX CC compositions containing CE proteins, are used, therapeutically or as
XX CC vaccines, to protect particularly mammals and birds, specifically cats
XX CC and dogs, against haematophagous ectoparasite infestation (HPI),
XX CC specifically fleas. More generally the compositions can be used to treat
XX CC arthropods generally, including pests of agricultural crops, trees,
XX CC stored goods etc., also those that are vectors of disease. Fragments of
XX CC the DNA can be used as probes and primers for identification or
XX CC production of nucleic acid. Antibodies against the protein can be used
XX CC for passive immunisation: to screen expression libraries; to isolate the
XX CC protein and to target cytotoxic compounds to HEP. The compounds
XX CC containing CE are effective against both adult and larval stages; they
XX CC target CEs, including juvenile hormone, that are involved in
XX CC development, metamorphosis, feeding, digestion and reproduction.
XX SO Sequence 1515 BP; 504 A; 268 C; 314 G; 428 T; 1 other.

Query Match 0.6%; Score 18; DB 19; Length 1515;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1002 ctccaattcagaatcc 1019
b 363 CTCCAATTTCAGAAATCC 346

RESULT 46
AAV40738
ID AAV40738 standard; cDNA; 1515 BP.
XX
XX AC AAV40738;
XX XX
XX DT 23-SEP-1998 (first entry)
XX XX
XX DE C. felis esterase, nfe51515, coding sequence complementary strand.
XX KW Esterase; flea; protective immune response; carboxylesterase; arthropod;
XX KW haematophagous ectoparasite infestation; nfe51515; ds.
XX XX
XX OS Ctenocephalides felis.
XX XX
XX PN WO9821324-A1.
XX PD 22-MAY-1998.
XX XX

```

```

PF 10-NOV-1997; 97WO-US20598.
XX XX
XX PR 12-NOV-1996; 96US-0747221.
XX XX
XX PA (HESK-) HESKA CORP.
XX XX
XX PI Brandt KS, Silver GM, Wisniewski N;
XX XX
XX DR WPI; 1998-297929/26.
XX XX
XX PT New nucleic acid encoding carboxyl:esterase(s) from fleas - useful,
XX PT e.g. in vaccines, for preventing infestation by haematophagous
XX PT ectoparasites, particularly on cats and dogs
XX XX
XX (S) Claim 1; Page 115-116; 230pp; English.
XX XX
XX CC This sequence is the complementary strand of the DNA encoding the flea
XX CC esterase protein, nfe51515 (see AAV40737 for coding strand), of the
XX CC invention. When administered to animals, the protein induces a
XX CC protective immune (antibody) response against carboxylesterase (CE), so
XX CC they, or compositions containing CE proteins, are used, therapeutically
XX CC or as vaccines, to protect particularly mammals and birds, specifically
XX CC cats and dogs, against haematophagous ectoparasite infestation (HPI),
XX CC specifically fleas. More generally the compositions can be used to treat
XX CC arthropods generally, including pests of agricultural crops, trees,
XX CC stored goods etc., also those that are vectors of disease. Fragments of
XX CC the DNA can be used as probes and primers for identification or
XX CC production of nucleic acid. Antibodies against the protein can be used
XX CC for passive immunisation: to screen expression libraries; to isolate the
XX CC protein and to target cytotoxic compounds to HEP. The compounds
XX CC containing CE are effective against both adult and larval stages; they
XX CC target CEs, including juvenile hormone, that are involved in
XX CC development, metamorphosis, feeding, digestion and reproduction.
XX SO Sequence 1515 BP; 428 A; 314 C; 268 G; 504 T; 1 other.

Query Match 0.6%; Score 18; DB 19; Length 1515;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1002 ctccaattcagaatcc 1019
Db 1153 ctccaattcagaatcc 1170

RESULT 47
AAF21032
ID AAF21032 standard; DNA; 1561 BP.
XX
XX AC AAF21032;
XX XX
XX DT 14-MAR-2001 (first entry)
XX XX
XX DE Human low adenosine antisense oligonucleotide related sequence #2399.
XX XX
XX KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
XX KW human; airway disorder; bronchoconstriction; lung inflammation;
XX KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
XX KW immunosuppressive; antiasmatic; analgesic; hypotensive; cytostatic;
XX KW respiratory obstruction; pulmonary obstruction; impeded respiration;
XX KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
XX KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
XX KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
XX KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
XX KW cancer; ss.
XX XX
XX OS Homo sapiens.
XX XX
XX PN WO200062736-A2.
XX PD 26-OCT-2000.
XX XX

```

PF 24-MAR-2000; 2000WO-US08020.
XX
PR 06-APR-1999; 99US-0127958.
XX
PA (UYEC-) UNIV EAST CAROLINA.
PA (NYCE/) NYCE J W.
XX
PI Nycce JW.
XX
DR WPI; 2000-679539/66.
XX
PT Low adenosine (A) content antisense oligonucleotides which do not
PT trigger receptors during metabolism, useful e.g. for treating
PT cancers and respiratory obstructions -
XX
PS Disclosure: Page 840; 1592pp; English.
XX
CC The present invention describes low adenosine (A) content antisense
CC oligonucleotides and compositions (I) comprising them. In the antisense
CC oligonucleotides the A is replaced by a 'universal' or alternative base.
CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC immunosuppressive, antisthmatic, hypotensive and cytostatic activities.
CC The antisense oligonucleotides and (I) can be used to down-regulate the
CC expression and/or activity of target polypeptides associated with
CC lung/respiratory disorders and malignancies, such as stimulating and
CC activating peptide factors and transmitters, transcription factors,
CC immunoglobulins and antibodies, antibody receptors, cytokines and
CC chemokines, endogenously produced specific and non-specific enzymes,
CC binding proteins, adhesion molecules and their receptors, cytokine and
CC chemokine receptors, adenosine receptors, bradykinin receptors, central
CC nervous system (CNS) and peripheral nervous and non-nervous system
CC receptors, CNS and peripheral nervous and non-nervous system peptide
CC transmitters, defensins, growth factors, vasoactive peptides and
CC antisense oligonucleotides may be used in this way to treat disorders
CC including respiratory obstruction (especially pulmonary obstruction
CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC and/or surfactant hypoproduction which are associated with a disease or
CC condition selected from pulmonary vasoconstriction, inflammation,
CC allergies, asthma, impeded respiration, respiratory distress syndrome
CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention.
XX
CC Sequence 1561 BP; 354 A; 445 C; 485 G; 277 T; 0 other;
XX
CC
XX
Query Match 0.6%; Score 18; DB 21; Length 1561;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1081 tgggtgaagtgaattcc 1098
DB 142 tgggtgaagtgaattcc 159
RESULT 48
AAA34910
ID AAA34910 standard; DNA; 1561 BP.
XX
AC AAA34910;
XX
DT 28-JUL-2000 (first entry)
XX
DE Human adenosine receptor related polynucleotide SEQ ID NO:2599.
XX
KW Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW phosphorothioate; impaired respiration; inflammation; allergy;
KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;

KW Lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS Homo sapiens.
XX
PN WO200009525-A2.
XX
PD 24-FEB-2000.
XX
PF 03-AUG-1999; 99WO-US17712.
XX
PR 03-AUG-1998; 98US-0095212.
XX
PA (UYEC-) UNIV EAST CAROLINA.
XX
PI Nycce JW.
XX
DR WPI; 2000-205971/18.
XX
CC New antisense oligonucleotides useful for treating e.g. pulmonary
CC vasoconstriction, inflammation, allergies, asthma, hypertension,
CC bronchitis, emphysema, respiratory distress syndrome, ischemia or
CC cancers -
XX
PS Disclosure: Page 769; 1343pp; English.
XX
CC The present invention describes a new composition comprising an
CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC inflammation. The ON can have antiinflammatory, antiallergic,
CC antisthmatic, cytostatic and analgesic activities. The compositions are
CC useful for the treatment of diseases associated with inflammation,
CC impaired airways, including lung disease and diseases whose secondary
CC effects afflict the lungs of a subject. They can be used for treating
CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
CC asthma, impeded respiration, respiratory distress syndrome, pain, cystic
CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC pulmonary disease (COPD), and cancers such as leukemias, lymphomas,
CC carcinomas, and cancers which may metastasize to the lungs, including
CC breast and prostate cancer. The reduction of the adenosine content of
CC the ONs reduces side effects. The A-containing ONs break down with the
CC release of deoxyadenosine which activates adenosine receptors causing
CC bronchoconstriction and inflammation. AAA32313 to AAA33312 represent the
CC nucleotide sequences given in the sequence listing from the present
CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC differ from the previously named sequences. SEQ ID NO:11 to 1680
CC (AAA32323 to AAA33992) are specifically claimed ONs from the present
CC invention. N.B. Sequences given in the disclosure of the present
CC invention do not match up with their corresponding SEQ ID NO: sequences
CC given in the sequence listing.
XX
XX
Sequence 1561 BP; 354 A; 445 C; 485 G; 277 T; 0 other;
XX
XX
Query Match 0.6%; Score 18; DB 21; Length 1561;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1081 tgggtgaagtgaattcc 1098
DB 142 tgggtgaagtgaattcc 159
RESULT 49
AAH13702/C
ID AAH13702 standard; cDNA; 1579 BP.
XX
AC AAH13702;
XX
DT 26-JUN-2001 (first entry)

```

XX DE Human cDNA sequence SEQ ID NO:10578.
XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX OS Homo sapiens.
XX PN EPI074617-A2.
XX PD 07-FEB-2001.
XX PE 28-JUL-2000; 2000EP-0116126.
XX PR 29-JUL-1999; 99JP-0248036.
XX PR 27-AUG-1999; 99JP-0300253.
XX PR 11-JAN-2000; 2000JP-0118776.
XX PR 02-MAY-2000; 2000JP-0183767.
XX PR 09-JUN-2000; 2000JP-0241899.
XX PA (HELI-) HELIX RES INST.
XX OT Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX IS Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WP: 2001-318749/34.
XX PT Primer sets for synthesizing polynucleotides, particularly the 5602
XX PT full-length cDNAs defined in the specification, and for the detection
XX PT and/or diagnosis of the abnormality of the proteins encoded by the
XX PT full-length cDNAs -
XX PS Claim 8; SEQ ID 10578; 2537bp + CD ROM; English.
XX CC The present invention describes primer sets for synthesizing 5602
XX CC full-length cDNAs defined in the specification, where a primer set
XX CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX CC to the complementary strand of a polynucleotide which comprises one of
XX CC the 5602 nucleotide sequences defined in the specification, where the
XX CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX CC of an oligonucleotide comprising a sequence complementary to the
XX CC complementary strand of a polynucleotide which comprises a 5'-end
XX CC sequence and an oligonucleotide comprising a sequence complementary to a
XX CC polynucleotide which comprises a 3'-end sequence, where the
XX CC oligonucleotide comprises at least 15 nucleotides and the combination of
XX CC the 5'-end sequence/3'-end sequence is selected from those defined in
XX CC the specification. The primer sets can be used in antisense therapy and
XX CC in gene therapy. The primers are useful for synthesizing polynucleotides,
XX CC particularly full-length cDNAs. The primers are also useful for the
XX CC detection and/or diagnosis of the abnormality of the proteins encoded by
XX CC the full-length cDNAs. The primers allow obtaining of the full-length
XX CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
XX CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX CC represent oligonucleotides, all of which are used in the exemplification
XX CC of the present invention.
XX SQ Sequence 1579 BP; 468 A; 274 C; 283 G; 554 T; 0 other.
XX QY
XX Query Match 0.6%; Score 18; DB 22; Length 1579;
XX Best Local Similarity 100.0%; Pred. No. 1.4e+02;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX QY 36 atgaattcataaaggac 53
XX ||||||||||||||||
XX DB 1229 ATGATTCATCAAGAGAC 1212
XX
XX RESULT 50
XX AAQ12528/C
XX ID AAQ12528 standard; DNA; 1611 BP.
XX AC . AAQ12528;

```

```

XX XX 25-SEP-1991 (first entry)
XX DT Thymidylate phosphorylase.
XX DE
XX KW Deoxyribonucleic acid; thymidine; dTMP; pyrimidine; ss.
XX KM
XX OS Bacillus subtilis CMG356 (ATCC 33234).
XX PN
XX PD Key Location/Qualifiers
XX FT 136..857
XX FT /*tag=
XX FT /product= dTMPase
XX FT CDS
XX PN WO9109130-A.
XX PD 27-JUN-1991.
XX PF 05-DEC-1990; 90WO-US06993.
XX PR 08-DEC-1989; 89US-0448158.
XX PA (CHEM-) CHEMGEN CORP.
XX PI Mc Danliss RL, Anderson DM;
XX DR WPI; 1991-208156/28.
XX DR P-PSDB; AAR12555.
XX PT Microorganism contg. deoxyribonucleic acid - encoding enzyme
XX PT causing accumulation of pyrimidine deoxyribonucleoside in
XX PT recoverable amts.
XX PS Disclosure: Fig 7(a-c); 79pp; English.
XX CC The sequence, comprising flanking DNA sequences obt'd. from a PSI
XX CC bacteriophage, encodes an enzyme capable of converting a thymidine
XX CC deoxyribonucleoside monophosphate to a thymidine deoxyribonucleoside.
XX CC The sequence may be used together with metabolic mutations of
XX CC heterologous DNA, encoding metabolic enzymes, to engineer cultured
XX CC cells to express thymidine deoxyribonucleosides (Tdms) in recoverable
XX CC amts., providing a fermentation source of Tdms.
XX CC See also AAQ12526-27.
XX SQ Sequence 1611 BP; 683 A; 171 C; 244 G; 513 T; 0 other;
XX QY
XX Query Match 0.6%; Score 18; DB 12; Length 1611;
XX Best Local Similarity 100.0%; Pred. No. 1.4e+02;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX QY 343 aggaattaaaggactgt 360
XX ||||||||||||||||
XX DB 979 AGGATTTAAGGACCTTGT 962
XX
XX RESULT 51
XX AAA47151/C
XX ID AAA47151 standard; DNA; 1611 BP.
XX AC AAA47151;
XX DT 03-OCT-2000 (first entry)
XX DE DNA encoding a serine protease inhibitor protein.
XX KW Serine protease inhibitor; green-lipped mussel; anti-thrombin;
XX KW divalent metal cation binding activity; dietary supplement;
XX KW anticoagulant; ss.
XX OS Perna canaliculus.
XX FH
XX Key Location/Qualifiers

```



```
FT CDS 1..1494
FT /*tag= a
FT /product= "serine protease inhibitor"
FT polyA_site 1557..1563
FT /*tag= b
PN WO200039165-A1.
XX
XX
XX
XX PD 06-JUL-2000.
XX
XX PF 23-DEC-1999; 99WO-NZ00227.
XX
XX PR 23-DEC-1998; 98NZ-0333568.
XX PR 23-JUL-1999; 99NZ-0336906.
XX
XX (HORT-) HORTICULTURE & FOOD RES INST NEW ZEALAND.
XX
XX PI Scotti PD, Dearing SC, Greenwood DR, Newcomb RD;
XX
XX DR WPI: 2000-452375/39.
XX P-PSDB: AAY93750.
XX
XX PS New Perna canaliculus serine protease inhibitor protein exhibiting
XX PT anti-thrombin activity and divalent metal cation binding activity,
XX PR useful as an anticoagulant agent and as a dietary supplement -
XX
XX PS Claim 11; Page 10-11; 44pp; English.
XX
XX CC The present sequence encodes a serine protease inhibitor
XX CC protein. The protein is isolated from the green-lipped mussel
XX CC (Perna canaliculus), and exhibits, inter alia, anti-thrombin activity
XX CC and divalent metal cation binding activity. The serine protease
XX CC inhibitor protein has a molecular weight of about 55 kilo Daltons.
XX CC The protein, and its fragments, are useful in medicaments, in food,
XX CC as dietary supplements or as bioremediation agents. In the dietary
XX CC supplements, the protein is associated with or bound to at least one
XX CC divalent cation (such as calcium, magnesium or zinc) of dietary
XX CC significance. The proteins or their fragments are also useful as
XX CC anticoagulant agents.
XX
XX SQ Sequence 1611 BP; 499 A; 348 C; 360 G; 402 T; 2 other;

Query Match 0.6%; Score 18; DB 21; Length 1611;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1051 tcttggtgcatcactt 1068
|||||
DB 767 TCTTGTGTCATCACTT 750

RESULT 52
AAAI5908/C
ID AAAI5908 standard; CDNA; 1619 BP.
XX
XX AC AAAI5908;
XX
XX DT 12-JUN-2000 (first entry)
XX
XX DE Human protein clone HP10195 full length coding sequence.
XX
XX KW Human protein: hydrophobic domain; nutritional source; haematopoiesis;
KW cytokine production; cell proliferation; cell differentiation;
KW immune deficiency; infectious disease; autoimmune disorder; asthma;
KW multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
KW allergic reaction; osteoporosis; osteoarthritis; periodontal disease;
KW nervous system disorder; Alzheimer's disease; Parkinson's disease;
KW Huntington's disease; liver fibrosis; lung fibrosis; reperfusion injury;
KW systemic cytokine damage; tissue differentiation; contraceptive; stroke;
KW coagulation disorder; myocardial infarction; inflammatory condition;
KW septic shock; sepsis; ischaemia; reperfusion injury; arthritis; tumour;
KW nephritis; therapy; ss.
```

```
XX
XX OS Homo sapiens.
XX
XX PN WO200005367-A2.
XX
XX PD 03-FEB-2000.
XX
XX PF 22-JUL-1999; 99WO-JP03929.
XX
XX PR 24-JUL-1998; 98JP-0208820.
XX PR 07-AUG-1998; 98JP-0224105.
XX PR 25-AUG-1998; 98JP-0238116.
XX PR 09-SEP-1998; 98JP-0254736.
XX PR 29-SEP-1998; 98JP-0275505.
XX
XX PA (SACA ) SACAMI CHEM RES CENT.
XX PA (PROT-) PROTEGENE INC.
XX
XX PI Kato S, Kimura T;
XX
XX DR WPI: 2000-182694/16.
XX DR P-PSDB: AAY94850.
XX
XX PS Novel human proteins having hydrophobic domains useful for treating
XX PT osteoporosis, Alzheimer's disease, Parkinson's disease, asthma,
XX PR multiple sclerosis, rheumatoid arthritis, cancer, anaemia, and stroke -
XX
XX PS Claim 4; Page 192-194; 351pp; English.
XX
XX CC This sequence encodes a human protein of the invention, which has
XX CC hydrophobic domains. The DNA sequences can be used as a probe or as a
XX CC genetic marker. The protein can also be used as a marker, and to identify
XX CC potential genetic disorders. The DNA and protein can also be used as
XX CC nutritional sources or supplements. The protein exhibits cytokine, cell
XX CC proliferation, cell differentiation activities and induces production of
XX CC other cytokines in certain cell populations. The protein also exhibits
XX CC immune stimulating or immune suppressing activity. It can be used in the
XX CC treatment of various immune deficiencies and disorders, and to treat
XX CC infectious diseases caused by viral, bacterial, fungal or other
XX CC infections. The protein is also used for treating autoimmune disorders
XX CC such as multiple sclerosis, systemic lupus erythematosus, and rheumatoid
XX CC arthritis. It is also useful in the treatment of allergic reactions and
XX CC conditions such as asthma, and in immune suppression after organ
XX CC transplantation. The protein is useful in regulation of haematopoiesis
XX CC and consequently in the treatment of myeloid or lymphoid cell
XX CC deficiencies. It is also used in compositions for tissue growth or
XX CC regeneration. The protein is also used in the treatment of osteoporosis
XX CC or osteoarthritis and in the treatment of periodontal disease and other
XX CC tooth repair processes. The protein is used in the treatment of nervous
XX CC system disorders such as Alzheimer's disease, Parkinson's disease, and
XX CC Huntington's disease. They are useful for protection or regeneration and
XX CC treatment of lung or liver fibrosis, reperfusion injury in various
XX CC tissues, and conditions resulting from systemic cytokine damage. They are
XX CC also used for promoting or inhibiting tissue differentiation. They are
XX CC also used as contraceptives since they exhibit activin or inhibin related
XX CC activities and as a fertility inducing therapeutic. They are used for
XX CC treating various coagulation disorders and in treatment and prevention of
XX CC conditions resulting from coagulation activities e.g. myocardial
XX CC infarction or stroke. They also acts as receptors, receptor ligands or
XX CC inhibitors or agonists of receptor/ligand interactions. They are used to
XX CC treat inflammatory conditions such as septic shock, sepsis, ischaemia
XX CC reperfusion injury, arthritis, and nephritis. They can be used to
XX CC prevent tumours.
XX
XX SQ Sequence 1619 BP; 444 A; 359 C; 363 G; 453 T; 0 other;

Query Match 0.6%; Score 18; DB 21; Length 1619;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

36 atgaattcataagac 53
|||||
```

Db 1277 ATGATTTCATTAAGAC 1260

RESULT 53
AAV40760/c
ID AAV40760 standard; cDNA; 1650 BP.

XX
AC AAV40760;
XX
XX 23-SEP-1998 (first entry)
XX
XX C. felis esterase, nFE51650, coding sequence.
DE
XX
XX Esterase; flea; protective immune response; carboxylesterase; arthropod;
KW haematophagous ectoparasite infestation; nFE51650; ds.
XX
OS Ctenocephalides felis.
XX
PN MO9821324-A1.
XX
PD 22-MAY-1998.
XX
XX 10-NOV-1997; 97WO-US20598.
PR 12-NOV-1996; 96US-0747221.
XX
XX (HESK-) HESKA CORP.
PI Brandt KS, Silver GM, Wisniewski N;
XX
XX WPI: 1998-297929/26.
DR P-PSDB; AAM57865.
XX
XX New nucleic acid encoding carboxyl:esterase(s) from fleas - useful,
PT e.g. in vaccines, for preventing infestation by haematophagous
XX ectoparasites, particularly on cats and dogs
XX
PS Claim 1; Page 188-191; 230pp; English.

XX
XX This sequence encodes the flea esterase protein, nFE51650 (the
CC complementary strand is shown in AAV40761), of the invention. When
CC administered to animals, the protein induces a protective immune
CC (antibody) response against carboxylesterase (CE), so they, or
CC compositions containing CE proteins, are used, therapeutically or as
CC vaccines, to protect particularly mammals and birds, specifically cats
CC and dogs, against haematophagous ectoparasite infestation (HEP),
CC specifically fleas. More generally the compositions can be used to treat
CC arthropods generally, including pests of agricultural crops, trees,
CC stored goods etc., also those that are vectors of disease. Fragments of
CC the DNA can be used as probes and primers for identification or
CC production of nucleic acid. Antibodies against the protein can be used
CC for passive immunisation; to screen expression libraries; to isolate the
CC protein and to target cytotoxic compounds to HEP. The compounds
CC containing CE are effective against both adult and larval stages; they
CC target CEs, including juvenile hormone, that are involved in
CC development, metamorphosis, feeding, digestion and reproduction.

XX
XX Sequence 1650 BP; 542 A; 284 C; 341 G; 482 T; 1 other:

Query Match 0.6%; Score 18; DB 19; Length 1650;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1002 ctccaattcagaatcc 1019
|||||
Db 498 CTCCAAAFTCGAATCC 481

RESULT 54
AAV40761
ID AAV40761 standard; cDNA; 1650 BP.
XX

AC AAV40761;
XX
XX 23-SEP-1998 (first entry)
XX
XX C. felis esterase, nFE51650, coding sequence complementary strand.
DE
XX
XX Esterase; flea; protective immune response; carboxylesterase; arthropod;
KW haematophagous ectoparasite infestation; nFE51650; ds.
XX
OS Ctenocephalides felis.
XX
PN MO9821324-A1.
XX
PD 22-MAY-1998.
XX
XX 10-NOV-1997; 97WO-US20598.
PF 12-NOV-1996; 96US-0747221.
PR
XX
XX (HESK-) HESKA CORP.
PA
PI Brandt KS, Silver GM, Wisniewski N;
XX
XX WPI: 1998-297929/26.
DR
XX
XX New nucleic acid encoding carboxyl:esterase(s) from fleas - useful,
PT e.g. in vaccines, for preventing infestation by haematophagous
XX ectoparasites, particularly on cats and dogs
XX
XX Claim 1; Page 191-192; 230pp; English.

XX
XX This sequence is the complementary strand of the DNA encoding the flea
CC esterase protein, nFE51650 (see AAV40760 for coding strand), of the
CC invention. When administered to animals, the protein induces a
CC protective immune (antibody) response against carboxylesterase (CE), so
CC they, or compositions containing CE proteins, are used, therapeutically
CC or as vaccines, to protect particularly mammals and birds, specifically
CC cats and dogs, against haematophagous ectoparasite infestation (HEP),
CC specifically fleas. More generally the compositions can be used to treat
CC arthropods generally, including pests of agricultural crops, trees,
CC stored goods etc., also those that are vectors of disease. Fragments of
CC the DNA can be used as probes and primers for identification or
CC production of nucleic acid. Antibodies against the protein can be used
CC for passive immunisation; to screen expression libraries; to isolate the
CC protein and to target cytotoxic compounds to HEP. The compounds
CC containing CE are effective against both adult and larval stages; they
CC target CEs, including juvenile hormone, that are involved in
CC development, metamorphosis, feeding, digestion and reproduction.

XX
XX Sequence 1650 BP; 482 A; 341 C; 284 G; 542 T; 1 other:

Query Match 0.6%; Score 18; DB 19; Length 1650;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1002 ctccaattcagaatcc 1019
|||||
Db 1153 ctccaattcagaatcc 1170

RESULT 55
AAV71287/c
ID AAV71287 standard; DNA; 1723 BP.
XX
XX AAV71287;
XX
XX 12-AUG-1999 (first entry)
XX
XX Human vesicular binding protein nucleotide sequence.
DE
XX
XX VSBP; fibroblast; Aplysia; VAMP; vesicle associated membrane protein;
KW synapobrevin binding protein; cancer; inflammation; ss.

XX	OS	Homo sapiens.	
XX	PH	Key	Location/Qualifiers
XX	FT	CDS	568..1278
XX	FT		/*tag= a
XX	FT		/product= "MVBP"
XX	PN	W09851797-A1.	
XX	PD	19-NOV-1998.	
XX	PF	15-MAY-1998;	98WO-US10225.
XX	PR	15-MAY-1997;	97US-0857213.
XX	PA	(INCYTE) INCYTE PHARM INC.	
XX	PI	Bandman O, Hawkins PR, Murry LE;	
XX	DR	WPI; 1999-045229/04.	
XX	DR	P-PSDB; AAM86813.	
XX	PT	Human vesicle binding protein - useful for the diagnosis, treatment	
XX	PT	and prevention of cell proliferation disorders and inflammation	
XX	PS	Claim 5; Fig 1A-D; 58pp; English.	
XX	CC	The present sequence represents a nucleic acid encoding a purified human	
XX	CC	vesicle binding protein (designated MVBP). This was obtained by standard	
XX	CC	cloning and screening procedure, where it was first isolated in incyte	
XX	CC	clone 148415 from a normal fibroblast library (FIBRNG701). A comparison	
XX	CC	of the MVBP with the Aplysia vesicle-associated membrane protein	
XX	CC	(VAMP)/synaptobrevin binding protein showed that they shared 44%	
XX	CC	identity. MVBP and its products can be used to diagnose, treat or prevent	
XX	CC	disorders of cell proliferation (e.g. cancer) or inflammation. Sequences	
XX	CC	complementary to MVBP can be used to detect MVBP in samples. Abs are used	
XX	CC	to diagnose conditions associated with MVBP or in assays to monitor	
XX	CC	patients being treated with MVBP or agonists/antagonists.	
XX	SO	Sequence 1723 BP; 548 A; 356 C; 325 G; 468 T; 26 other;	
XX	Query Match	0.6%; Score 18; DB 20; Length 1723;	
XX	Best Local Similarity	100.0%; Pred. No. 1.4e+02;	
XX	Matches 18; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
XX	36 atgaattcattaaaggac 53		
XX			
XX	Db 1540 ATGAATTTCATTAAGGAC 1523		
XX	RESULT 56		
XX	AA067601		
XX	ID AA067601 standard; CDNA; 1800 BP.		
XX	AC AA067601;		
XX	DF 20-NOV-1994 (first entry)		
XX	DE Retinoblastoma protein Ap4.		
XX	Retinoblastoma-associated polypeptide: RAP; Ap12; Ap4;		
XX	transcription factor; FR2; cell cycle; ss.		
XX	OS Homo sapiens.		
XX	PN W09412521-A.		
XX	PD 09-JUN-1994.		
XX	PF 19-NOV-1993; 93WO-US11310.		

PR	20-NOV-1992;	.92US-0979156.
XX		
PA	(TEXA) UNIV TEXAS SYSTEM.	
XX		
PI	Lee W, Shan B;	
XX		
DR	WPI: 1994-200185/24.	
XX		
PT	Nucleic acid encoding retinoblastoma-associated polypeptide(s) -	
PR	used for producing prods. for use in studying cell cycling and	
PT	disregulated cell growth.	
XX		
PS	Disclosure: Page 37-38; 77pp; English.	
XX		
CC	CDNA encoding a retinoblastoma (RB) associated protein (RAP), Ap12,	
CC	was cloned by direct screening of cDNA expression libraries using	
CC	purified RB protein as probe. Ap12 showed transcription factor E2F	
CC	and RB binding activities. The 5' and 3' cDNA sequences were	
CC	determined for Ap2 (AA067395, AA067596), Ap8 (AA067598) and	
CC	Ap13 (AA067599, AA067600), and full-length encoding sequences for Ap4	
CC	(AA067601) and Ap10 (AA067601), which also encoded RB-associated	
CC	proteins.	
XX		
SQ	Sequence 1800 BP; 747 A; 333 C; 372 G; 348 T; 0 other;	
	Query Match	0.6%; Score 18; DB 15; Length 1800;
	Best Local Similarity	100.0%; Pred. No. 1.4e+02;
	Matches	18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	2831 aaagaaccctcgaaaaa 2848	
Db	307 aaagaaccctcgaaaaa 324	
RESULT 57		
AAF25374		
ID	AAF25374 standard; CDNA; 1878 BP.	
XX		
AC	AAF25374;	
XX		
DT	15-MAY-2001 (first entry)	
DE		
XX	Nucleotide sequence of the Arabidopsis SGS3 polypeptide.	
XX		
KM	SGS3 gene: post-transcriptional inactivation; RNA degradation;	
KW	viral resistance; resistance; fatty acid content; protein content; ss.	
XX		
OS	Arabidopsis thaliana.	
XX		
XX		
FH	Key	Location/Qualifiers
FT	CDS	1..1878
FT		/*tag= a
FT		/product= "SGS3"
PN	WO200105951-A2.	
XX		
PD	25-JAN-2001.	
XX		
PF	13-JUL-2000; 2000WO-FR02052.	
XX		
PR	16-JUL-1999; 99FR-0009417.	
PR	26-JAN-2000; 2000FR-0001006.	
XX		
PA	(AVERT) AVENTIS CROSCIENCE SA.	
XX	(INRG) INST NAT RECH AGRONOMIQUE.	
PI	Becjin C, Elmayan T, Vaucheret H;	
XX		
DR	WPI: 2001-159529/16.	
XX	P-PEDB; AAB31798.	
XX		
PT	New SGS3 gene from Arabidopsis thaliana, useful for increasing virus	

PT resistance in plants and, when inhibited, for increasing transgene
 expression -
 XX
 PS Claim 1; Page 32-35; 36pp; French.
 XX
 CC The present sequence encodes the Arabidopsis thaliana SGS3 polypeptide.
 CC (The SGS3 gene is essential for post-transcriptional inactivation
 CC (degradation of RNA) and for resistance to viruses. Overexpression
 CC of SGS3 results in plants with increased resistance to viruses,
 CC while inactivation of SGS3 in transgenic plants (e.g. by expressing
 CC antisense RNA, by mutation or by homologous recombination) increases
 CC the level of the transgene product. This product may e.g. impart
 CC resistance (to herbicide, insects or pathogens), alter contents
 CC of essential fatty acids or proteins, or is pharmaceutically active,
 CC e.g. an immunoglobulin or interferon.
 CC
 XX Sequence 1878 BP; 575 A; 332 C; 554 G; 417 T; 0 other:
 SQ

Query Match 0.6%; Score 18; DB 22; Length 1878;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 1020 aggtgcttgaggaaatc 1037
 ||||||||||||||||
 Db 1388 aggtgcttgaggaaatc 1405

RESULT 58
 AAV82522
 ID AAV82522 standard; DNA; 1926 BP.
 XX
 AC AAV82522;
 XX
 DT 17-MAR-1999 (first entry)
 XX
 DE Sphingomonas capsulata mature aminopeptidase I encoding DNA.
 XX
 KW Aminopeptidase; protein hydrolysate; glycine releasing; protease;
 KW proteolaceous material; flavour; food; baking; animal feed additive;
 KW palatability; hydrolysis; solubility; emulsifying; foaming; aroma;
 KW mouthfeel; crust colour; ss.
 XX
 OS Sphingomonas capsulata.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1491
 FT /*tag= a
 XX
 PN W09851163-A2.
 XX
 PR 19-NOV-1998.
 XX
 PF 15-MAY-1998; 98WO-US09998.
 XX
 PR 16-DEC-1997; 97US-0069719.
 PR 16-MAY-1997; 97US-0857886.
 PR 20-OCT-1997; 97US-0062893.
 PR 16-DEC-1997; 97DK-0001465.
 XX
 PA (ASAHI) ASahi CHEM. IND CO. LTD.
 PA (NOVO) NOVO NORDISK BIOTECH INC.
 PA (NOVO) NOVO-NORDISK AS.
 XX
 PI Blinkovsky A, Brown K, Byun T, Fujii M, Golightly E;
 PI Kotod IV, Marumotac, Mathiansen TE;
 XX
 DR WPI: 1999-045177/04.
 DR P-PSDB: AAW89587.
 XX
 PT Production of protein hydrolysate - using protease and enzyme that
 PT releases glycine, useful as flavour improvers in foods and animal
 PT feed additives

XX
 PS Claim 7; Page 61; 84pp; English.
 XX
 CC A method has been developed for the production of protein hydrolysates
 CC (PH) comprising reacting a protein with: (i) at least one polypeptide
 CC with Gly-releasing activity; and (ii) at least one other protease so
 CC that the amount of Gly produced is greater than when the protease is
 CC used alone. PH are used to improve flavour of foods (e.g. baked goods)
 CC and as animal feed additives. PH are preferably also enriched in Glu
 CC (free and/or peptide bound), so have improved flavour and palatability.
 CC Addition of a polypeptide with Gly-releasing activity increases the
 CC degree of hydrolysis (or reduces the amount of enzyme needed) and
 CC hydrolysates have better solubility, and emulsifying and foaming
 CC properties. Baked goods containing them have improved aroma, mouthfeel
 CC and crust colour. The present sequence encodes Sphingomonas capsulata
 CC aminopeptidase I, which is used in the method of the invention.
 CC
 XX Sequence 1926 BP; 362 A; 633 C; 628 G; 303 T; 0 other:
 SQ

Query Match 0.6%; Score 18; DB 20; Length 1926;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 2127 gctggtgtgtgctggaac 2144
 ||||||||||||||||
 Db 1242 gctggtgtgtgctggaac 1259

RESULT 59
 AAZ77504
 ID AAZ77504 standard; CDNA; 1962 BP.
 XX
 AC AAZ77504;
 XX
 DT 10-APR-2000 (first entry)
 XX
 DE Human ovarian tumor CDNA library derived EST fragment 55.
 XX
 KW Expressed sequence tag; EST; human; ovarian tumor; anticancer;
 KW gene therapy; treatment; ss.
 XX
 OS Homo sapiens.
 XX
 PN DE19817557-A1.
 XX
 PD 21-OCT-1999.
 XX
 PF 09-APR-1998; 98DE-1017557.
 XX
 PR 09-APR-1998; 98DE-1017557.
 XX
 PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
 XX
 PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
 XX
 DR WPI: 1999-591920/51.
 DR P-PSDB: AAY76633, AAY76634, AAY76635.
 XX
 PT New nucleic acid sequences expressed in ovarian, and some other, cancer
 PT tissues, and derived polypeptides, for treatment of ovarian cancer and
 PT identification of therapeutic agents -
 XX
 PS Claim 3; Page 182-183; 310pp; German.
 XX
 CC This invention describes novel nucleic acid (cDNA) sequences (A) which
 CC have anticancer activity and are highly expressed in ovarian tumor
 CC tissue (and some also in testis and breast cancer tissue). The products
 CC of the invention can be used for gene therapy. (A) are used (i) for
 CC recombinant expression of polypeptides (B) and (ii) to isolate complete
 CC genes. (B) are used (i) to identify agents suitable for treatment of
 CC ovarian cancer; (ii) directly for treating this form of cancer
 CC (including expression from gene therapy vectors) and (iii) for generation

CC of specific antibodies. (A) are identified by assembling ESTs (expressed
 CC sequence tags) from a particular tissue type before comparison of
 CC expression patterns. This allows a significantly longer fragment of the
 CC gene to be revealed, so should reduce the number of failures associated
 CC with the fact that ESTs from different libraries may represent different
 CC parts of the same unknown gene, distorting the estimated frequency of
 CC occurrence in a particular tissue. AAZ77450-277572 represent the human
 CC ovarian tumor cDNA library derived EST fragments described in the method
 CC of the invention and encode the protein fragments represented in
 CC AAV76505-Y76638.

XX Sequence 1962 BP; 596 A; 357 C; 389 G; 620 T; 0 other;

Query Match

Best Local Similarity 100.0%; Score 18; DB 20; Length 1962;
 Pred. No. 1.4e+02;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 36 atgaattcataagac 53
 |||
 Db 368 atgaattcataagac 385

RESULT 60

AAV40735/C
 ID AAV40735 standard; cDNA; 1982 BP.

XX AAV40735;

DT 23-SEP-1998 (first entry)

XX C. felis esterase, nFE51982, coding sequence.

KM Esterase; flea; protective immune response; carboxylesterase; arthropod;
 KM haematophagous ectoparasite infestation; nFE51982; ds.

OS Ctenocephalides felis.

XX Key Location/Qualifiers
 FT CDS 31..1520
 FT /*tag= a

PN WO9821324-A1.

PD 22-MAY-1998.

PF 10-NOV-1997; 97WO-US20598.

PR 12-NOV-1996; 96US-0747221.

XX (HESK-) HESKA CORP.

PI Brandt KS, Silver GM, Wisniewski N;

DR WPI; 1998-297929/26.

XX P-SDB; AAM57851.

PT New nucleic acid encoding carboxylesterase(s) from fleas - useful,
 PT e.g. in vaccines, for preventing infestation by haematophagous
 PT ectoparasites, particularly on cats and dogs

PS Claim 1; Page 106-109; 230pp; English.

XX This sequence encodes the flea esterase protein, nFE51982 (the
 CC complementary strand is shown in AAV40736), of the invention. When
 CC administered to animals, the protein induces a protective immune
 CC (antibody) response against carboxylesterase (CE), so they, or
 CC compositions containing CE proteins, are used, therapeutically or as
 CC vaccines, to protect particularly mammals and birds, specifically cats
 CC and dogs, against haematophagous ectoparasite infestation (HEP),
 CC specifically fleas. More generally the compositions can be used to treat
 CC arthropods generally, including pests of agricultural crops, trees,
 CC stored goods etc., also those that are vectors of disease. Fragments of

CC the DNA can be used as probes and primers for identification or
 CC production of nucleic acid. Antibodies against the protein can be used
 CC for passive immunisation; to screen expression libraries; to isolate the
 CC protein and to target cytotoxic compounds to HEP. The compounds
 CC containing CE are effective against both adult and larval stages; they
 CC target CEs, including juvenile hormone, that are involved in
 CC development, metamorphosis, feeding, digestion and reproduction.

XX Sequence 1982 BP; 691 A; 304 C; 357 G; 629 T; 1 other;

Query Match

Best Local Similarity 100.0%; Score 18; DB 19; Length 1982;
 Pred. No. 1.4e+02;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1002 ctccaattcagaatcc 1019
 |||
 Db 365 CTCCAATTTCAGAAATCC 348

RESULT 61

AAV40736
 ID AAV40736 standard; cDNA; 1982 BP.

XX AAV40736;

DT 23-SEP-1998 (first entry)

XX C. felis esterase, nFE51982, coding sequence complementary strand.

KM Esterase; flea; protective immune response; carboxylesterase; arthropod;
 KM haematophagous ectoparasite infestation; nFE51982; ds.

XX Ctenocephalides felis.

PN WO9821324-A1.

PD 22-MAY-1998.

PF 10-NOV-1997; 97WO-US20598.

PR 12-NOV-1996; 96US-0747221.

XX (HESK-) HESKA CORP.

PI Brandt KS, Silver GM, Wisniewski N;

DR WPI; 1998-297929/26.

PT New nucleic acid encoding carboxylesterase(s) from fleas - useful,
 PT e.g. in vaccines, for preventing infestation by haematophagous
 PT ectoparasites, particularly on cats and dogs

PS Claim 1; Page 111-112; 230pp; English.

XX This sequence is the complementary strand of the DNA encoding the flea
 CC esterase protein, nFE51982 (see AAV40735 for coding strand), of the
 CC invention. When administered to animals, the protein induces a
 CC protective immune (antibody) response against carboxylesterase (CE), so
 CC they, or compositions containing CE proteins, are used, therapeutically
 CC or as vaccines, to protect particularly mammals and birds, specifically
 CC cats and dogs, against haematophagous ectoparasite infestation (HEP),
 CC specifically fleas. More generally the compositions can be used to treat
 CC arthropods generally, including pests of agricultural crops, trees,
 CC stored goods etc., also those that are vectors of disease. Fragments of
 CC the DNA can be used as probes and primers for identification or
 CC production of nucleic acid. Antibodies against the protein can be used
 CC for passive immunisation; to screen expression libraries; to isolate the
 CC protein and to target cytotoxic compounds to HEP. The compounds
 CC containing CE are effective against both adult and larval stages; they
 CC target CEs, including juvenile hormone, that are involved in
 CC development, metamorphosis, feeding, digestion and reproduction.

RESULT 64
AAZ96326/c
ID AAZ96326 standard; DNA: 2299 BP.
XX
AC AAZ96326;
XX
DT 10-APR-2000 (first entry)
XX
DE S. pneumoniae derived DNA from ORF #154.
XX
KW Treatment; prevention; disease; diagnosis; gene therapy; screening;
KW bacterial; antimicrobial; antibiotic; pathogenesis; infection; ss.
XX
OS Streptococcus pneumoniae.
XX
PN WO9806734-A1.
XX
PD 19-FEB-1998.
XX
PF 15-AUG-1997; 97WO-US14436.
XX
PR 16-AUG-1996; 96US-0024022.
XX
PA (SMIR) SMITHKLINE BEECHAM CORP.
XX
PI Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;
PI Stodola RK;
XX
DR WPI: 1998-159452/14.
DR P-PSDB; AAY85980, AAY85981.
XX
PT Streptococcus pneumoniae proteins and related DNA - useful for
PT screening compounds for antibacterial activity
XX
PS Claim 4; Page 188-189; 640pp; English.
XX
XS This invention describes novel isolated Streptococcus pneumoniae
CC polynucleotides (see AAZ96173-296494) and their encoded proteins (see
CC AAY87297-186182). The DNA, vectors and host cells described in the
CC method of the invention are useful for the recombinant expression of the
CC polypeptides. The polypeptides are useful for treatment or prevention of
CC disease, or diagnosis of disease related to expression or activity of
CC such a polypeptide. They can also be used to screen for compounds which
CC interact with and inhibit or activate such a polypeptide. The
CC polypeptides (or DNA encoding them, via gene therapy) are also useful
CC for inducing an immunological response in a mammal. The antagonists are
CC useful to inhibit such bacterial polypeptides. The polypeptides are
CC particularly useful to identify antimicrobial compounds and antibiotics.
CC They are also useful to determine their role in pathogenesis of
CC infection, dysfunction and disease.
XX
SQ Sequence 2299 BP; 585 A; 514 C; 492 G; 708 T; 0 other;

Query Match 0.6%; Score 18; DB 19; Length 2299;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2765 caagcttggtgaaana 2782
|||||
DB 429 CAAGCTTGCGTGAANA 412

RESULT 65
AAV42995/c
ID AAV42995 standard; DNA: 2299 BP.
XX
AC AAV42995;
XX
DT 09-NOV-1998 (first entry)
XX
DE Streptococcus pneumoniae polypeptide coding region.

XX
KM Polypeptide; ORF; open reading frame; infection; bacterial;
KM streptococcal; bacteremia; diagnosis; prophylaxis; ds.
XX
OS Streptococcus pneumoniae.
XX
FH Key Location/Qualifiers
FT CDS complement (388..1008)
FT /*tag= a
FT /note= "polypeptide"
XX
PN WO9823631-A1.
XX
PD 04-JUN-1998.
XX
PF 24-NOV-1997; 97WO-US21976.
XX
PR 27-NOV-1996; 96US-0031879.
XX
PA (SMIR) SMITHKLINE BEECHAM CORP.
PA (SMIR) SMITHKLINE BEECHAM PLC.
XX
PI Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;
PI Reid RH, Zarfos PN;
XX
DR WPI: 1998-322654/28.
DR P-PSDB; AAW62719.
XX
XS Streptococcus pneumoniae polynucleotides - useful for developing
PT products for diagnosis, prevention and treatment of infections e.g.
PT pneumonia, bacteremia, meningitis or endocarditis
XX
PS Claim 1; Page 106; 181pp; English.
XX
XS The sequence is that of a Streptococcal polypeptide coding region.
CC The polypeptide can potentially be used for the diagnosis and
CC prevention of bacterial infections, especially SP infection.
CC It may be used for the treatment of diseases such as otitis media,
CC conjunctivitis, pneumonia, bacteremia, meningitis, sinusitis, pleural
CC empyema, endocarditis or infection of the cerebrospinal fluid.
XX
SQ Sequence 2299 BP; 585 A; 514 C; 491 G; 708 T; 1 other;

Query Match 0.6%; Score 18; DB 19; Length 2299;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2765 caagcttggtgaaana 2782
|||||
DB 429 CAAGCTTGCGTGAANA 412

RESULT 66
AAW84399
ID AAW84399 standard; DNA: 3000 BP.
XX
AC AAW84399;
XX
DT 09-SEP-1999 (first entry)
XX
DE S. capsulata IF012533 aminopeptidase coding sequence.
XX
KW Aminopeptidase; amino acid removal; protein hydrolysatate production;
KW hydrolysis; flavour development; enzyme deactivation;
KW peptide sequence cleavage; post-translational processing;
KW precursor protein activation; ss.
XX
OS Sphingomonas capsulata.
XX
PN WO931226-A1.
XX
PD 24-JUN-1999.

```

XX PF 13-NOV-1998; 98WO-DK00495.
XX PR 15-MAY-1998; 98DK-0000670.
PR 16-DEC-1997; 97DK-0001465.
PR 16-DEC-1997; 97US-0069719.
XX PA (ASAH ) ASAMI CHEM IND CO LTD.
PA (NOVO ) NOVO NORDISK BIOTECH INC.
XX PA (NOVO ) NOVO-NORDISK AS.
XX PI Blinkovsky A, Brown K, Byun TS, Fujii M, Klotz AV;
PI Matumoto C, Sioma A, Tang M;
XX DR WPI: 1999-418769/35.
DR P-PSDB: AAY22184.
XX PT New isolated aminopeptidase polypeptides used in, e.g. food industry
PT Claim 14; Fig 9; 84pp; English.
XX PS
XX CC This sequence encodes the Sphingomonas capsulata IF012533 aminopeptidase
CC of the invention. The aminopeptidase polypeptides catalyse the removal of
CC amino acids from the N-terminal end of peptides, oligopeptides or
CC proteins. They can be used in the production of protein hydrolysates for
CC enhancing the degree of hydrolysis and flavour development, particularly
CC in foods. They can also be used to deactivate enzymes. They can also be
CC used for specific cleavage of peptide sequences, e.g. to provide the
CC necessary post-translational processing to activate precursor proteins.
XX SO Sequence 3000 BP; 528 A; 978 C; 978 G; 516 T; 0 other;

Query Match 0.6%; Score 18; DB 20; Length 3000;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2127 gctggtgtgctggaagc 2144
DB 1911 gctggtgtgctggaagc 1928

RESULT 67
AAH51776/C
ID AAH51776 standard; DNA: 3001 BP.
XX AC AAH51776;
XX DT 29-AUG-2001 (first entry)
XX CC Chromosome 13q31-q33 biallelic marker containing amplicon SEQ ID 188.
XX PA sbg1; g34665; sbg2; g35017; g35018; chromosome 13q31-q33; haplotype;
XX KM biallelic marker; polymorphism; schizophrenia; bipolar disorder; ds.
XX OS Homo sapiens.
XX PN WO200058510-A2.
XX PD 05-OCT-2000.
XX PF 30-MAR-2000; 2000WO-IB00435.
XX PR 30-MAR-1999; 99US-0126903.
PR 30-APR-1999; 99US-0131971.
PR 30-APR-1999; 99US-0132065.
PR 14-JUL-1999; 99US-0143928.
PR 27-JUL-1999; 99US-0145915.
PR 29-JUL-1999; 99US-0146452.
PR 29-JUL-1999; 99US-0146453.
PR 28-OCT-1999; 99US-0162288.
XX PA (GEST ) GENSET.

```

```

XX PI Cohen D, Blumenfeld M, Chumakov I, Bougueleret L, Bihain B;
PI Essioux L;
XX DR WPI: 2000-619082/59.
XX PT Polynucleotides comprising sequences from sbg1 and g35018 biallelic
XX markers are used for genotyping and detecting schizophrenia or bipolar
XX disorder and predisposition to these disorders.
XX PS Claim 2; Page 680-681; 737pp; English.
XX CC AAH51601 represents a human genomic nucleotide sequence comprising sbg1,
CC g34665, sbg2, g35017 and g35018 nucleic acid sequences located on the
CC human chromosome 13q31-q33 locus. The nucleotide sequences contain
CC biallelic markers and polymorphisms. Sequences AAH51602 - AAH51626 and
CC AAH62907 - AAH62915 represent cDNA human sbg1 cDNA sequences and protein
CC products. AAH51627 - AAH51631 and AAH62916 - AAH62918 represent g35018
CC cDNA sequences and protein products. Primers AAH51632 - AAH51699 are used
CC to isolate sbg1 cDNAs, while sbg1 exons from different primates are
CC represented by sequences AAH51642 - AAH51699. Nucleotide sequences of
CC amplicons which comprise biallelic markers located on the chromosome
CC 13q31-q33 locus are represented in AAH51700 - AAH51817. Biallelic markers
CC are represented in the sequences by degenerate/undefined base codes. PCR
CC primers AAH51818 and AAH51819 are used in the isolation of sequences of
CC the invention. The biallelic marker containing nucleotide sequences are
CC used to determine the identity of the nucleotide at a biallelic marker in
CC a sample DNA sequence. The nucleotide sequences may be labelled and used
CC for genotyping by determining the identity of a nucleotide at a region
CC D-related biallelic marker in a biological sample from single or multiple
CC subjects. By determining the frequency of a biallelic marker in a
CC population an association between a genotype and a trait, a haplotype and
CC a trait and a phenotype and a trait can be detected. The sequences can be
CC used to determine a predisposition to or early onset of schizophrenia or
CC bipolar disorder or a beneficial response to or side effects related to
CC treatment against schizophrenia or bipolar disorder.
XX SQ Sequence 3001 BP; 993 A; 463 C; 533 G; 1011 T; 1 other;

Query Match 0.6%; Score 18; DB 21; Length 3001;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1663 aagaatctgaaagcca 1680
DB 181 AAGAAATCTGAAAGCCA 164

RESULT 68
AAH51787/C
ID AAH51787 standard; DNA: 3001 BP.
XX AC AAH51787;
XX DT 29-AUG-2001 (first entry)
XX CC Chromosome 13q31-q33 biallelic marker containing amplicon SEQ ID 199.
XX PA sbg1; g34665; sbg2; g35017; g35018; chromosome 13q31-q33; haplotype;
XX KM biallelic marker; polymorphism; schizophrenia; bipolar disorder; ds.
XX OS Homo sapiens.
XX PN WO200058510-A2.
XX PD 05-OCT-2000.
XX PF 30-MAR-2000; 2000WO-IB00435.
XX PR 30-MAR-1999; 99US-0126903.
PR 30-APR-1999; 99US-0131971.
PR 30-APR-1999; 99US-0132065.

```


KW	Human; diagnosis; colon cancer; cancer; malignant; chromosome mapping;
KV	detection; colon cancer cell line Km12L4-A; ss.
XX	
OS	Homo sapiens.
PN	MO200018916-A2.
PD	06-APR-2000.
PE	23-SEP-1999; 99WO-US22226.
PR	28-SEP-1998; 98US-0102161.
PR	28-SEP-1998; 98US-0102180.
PR	29-SEP-1998; 98US-0102380.
PR	08-OCT-1998; 98US-0103815.
PR	27-OCT-1998; 98US-0105877.
PA	(CHIR) CHIRON CORP.
PA	(HYSE-) HYSEQ INC.
PI	Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PT	Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
PT	Lanson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
PS	Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
XX	
XX	WPI: 2000-293155/25.
XX	
XX	Polynucleotide library comprising 1079 defined sequences, useful in
PT	the form of an array to detect cancer or susceptibility to cancer -
PT	Claim 1; Page 501-502; 502pp; English.
CC	The present invention describes a library of polynucleotides comprising
CC	1079 nucleotide sequences (given in AAH30067 to AAH31145). Also described
CC	are: (1) an isolated polynucleotide (I) having at least 90% identity to
CC	one of the 1079 sequences; (2) a recombinant host cell containing (I);
CC	(3) an isolated polypeptide (II) encoded by (I); (4) an antibody that
CC	specifically binds to (II); (5) a vector comprising (I); and (6) a method
CC	of detecting differentially expressed genes correlated with a cancerous
CC	state of a mammalian cell comprising detecting a gene product encoded by
CC	65 of the 1079 sequences given in the specification. The polynucleotides
CC	are used to monitor patients having (or susceptible) to cancer to detect
CC	potentially malignant events at a molecular level before they are
CC	detectable at a gross morphological level. The polynucleotides are also
CC	useful for monitoring the efficacy of various therapies and preventive
CC	interventions. Polynucleotide probes based on the disclosed sequences
CC	are useful for chromosome mapping and detection of transcription levels.
CC	The 1079 polynucleotide sequences were derived from a human colon cancer
CC	cell line Km12L4-A CDNA library.
XX	
XX	Sequence 3529 BP: 1248 A; 778 C; 692 G; 803 T; 8 other:
QY	
DB	3465 aaagaaccctcgtgaaaa 3482
QY	2831 aaagacctctgaataa 2848
DB	
QY	18; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
XX	
XX	Query Match 0.6%; Score 18; DB 21; Length 3529;
XX	Best Local Similarity 100.0%; Pred. No. 1.5e+02;
XX	Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
XX	
XX	RESULT 71
XX	AAVB0604/C
XX	ID AAVB0604 standard; cDNA: 3942 BP.
XX	
XX	AAVB0604;
XX	
XX	01-MAR-1999 (first entry)
XX	
XX	Kidney injury associated molecule HM050 CDNA clone.
XX	
XX	Kidney injury associated molecule; kidney injury related molecule;
XX	

KW	KIM; tissue growth promotion; regeneration; renal condition;
KM	acute renal failure; acute nephritis; tumour; ds.
XX	Rattus sp.
OS	
FH	Key Location/Qualifiers
FT	CDS 7..1329
FT	/tag= a
FT	/product= "kidney injury associated molecule"
FT	complement (3542..3733)
FT	/tag= b
FT	/label= SAC_24644
XX	
PN	WO9853071-A1.
PD	26-NOV-1998.
XX	
PJ	22-MAY-1998; 98WO-US10547.
PR	23-MAY-1997; 97US-0047491.
PR	23-MAY-1997; 97US-0047490.
PA	(BIOJ) BIOGEN INC.
PI	Cate RL, Hession CA, Sanicola-Nadel M, Wei H;
DR	WPI; 1999-045312/04.
DR	P-PsDB; AAW86320.
XX	
PT	Kidney injury-associated molecule, KIM, polypeptides - upregulated
PT	in injured or regenerating tissues, useful to promote tissue growth
PT	and regeneration, especially to treat renal conditions
XX	
PS	Claim 9; Page 109-113; 213pp; English.
CC	The present sequence represents a kidney injury associated molecule
CC	(KIM) cDNA clone. KIM proteins can be administered therapeutically
CC	by expressing KIM encoding polynucleotides, to promote growth and/or
CC	survival of damaged tissue (e.g. renal tissue), since the KIM proteins
CC	are upregulated in injured or regenerating (especially renal) tissues.
CC	KIM fusion proteins, conjugates, antibodies and vectors can also be used
CC	therapeutically, e.g. these or the KIM proteins may be included with an
CC	acceptable carrier in pharmaceutical compositions, useful for therapy of
CC	prophylaxis or conditions associated with dysfunction/dysregulation of
CC	KIM genes or proteins, especially renal diseases or impairments of renal
CC	function in humans (e.g. acute renal failure, acute nephritis). The
CC	polynucleotides can be used to produce antisense sequences which, when
CC	internalised into cells, can disrupt expression of a cellular KIM gene,
CC	also useful in therapy (e.g. to block the growth of tumours dependent on
CC	KIM for growth) or compositions. The proteins and polynucleotides are
CC	useful diagnostically e.g. to detect and quantify renal injury/disease
CC	(indicative of increased risk, or presence of, renal injury or impaired
CC	function), or abnormal responses to tissue injury (indicative of
CC	increased risk, or presence of, an autoimmune response or abnormal
CC	tissue growth arising from/affecting renal tissue). The proteins can
CC	also be used to locate KIM-producing cells (especially specific loci,
CC	e.g. tissue masses abnormally producing/expressing KIM such as tumours
CC	arising from/affecting renal tissue), by contacting cells with an
CC	imageable KIM-binding reagent and imaging reagent accumulation.
XX	
SQ	Sequence 3942 BP; 908 A; 1019 C; 1087 G; 928 T; 0 other:
Query Match	0.6%; Score 18; DB 20; Length 3942;
Best Local Similarity	100.0%; Pred. No. 1.5e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY 1918 cagaagacacagtgcga 1935	
Db 2495 CAGAAGACACAGCTGGA 2478	

ID	AA064473/c	standard; DNA: 3969 BP.
XX	AA064473;	
XX	02-JAN-2001	(first entry)
XX	DNA encoding an astaxanthin synthetase polypeptide.	
XX	Astaxanthin synthetase; astaxanthin; beta-carotene; carotenogenic yeast;	
XX	antioxidant; cancer; colouring reagent; farmed fish; salmon; ss.	
XX	Phaffia rhodozyma.	
XX	Key	Location/Qualifiers
XX	CDS	535..3700
XX	5'UTR	517..518
XX	exon	535..652
XX	intron	653..734
XX	exon	735..783
XX	intron	784..898
XX	exon	899..1015
XX	intron	1016..1087
XX	exon	1088..1179
XX	intron	1180..1302
XX	exon	1303..1517
XX	intron	1518..1600
XX	exon	1601..1634
XX	intron	1635..1723
XX	exon	1724..1866
XX	intron	1867..1939
XX	exon	1940..1999
XX	intron	2000..2081
XX	exon	2082..2181
XX	intron	2182..2257
XX	exon	2258..2357

FT	exon	2258..2354
FT		/*tag= u
FT		/number=10
FT	Intron	2335..2431
FT		/*tag= v
FT		/number=10
FT	exon	2432..2542
FT		/*tag= w
FT		/number=11
FT	Intron	2543..2618
FT		/*tag= x
FT		/number=11
FT	exon	2619..2652
FT		/*tag= y
FT		/number=12
FT	Intron	2653..2742
FT		/*tag= z
FT		/number=12
FT	exon	2743..2814
FT		/*tag= aa
FT		/number=13
FT	Intron	2815..2962
FT		/*tag= ab
FT		/number=13
FT	exon	2963..3050
FT		/*tag= ac
FT		/number=14
FT	Intron	3051..3113
FT		/*tag= ad
FT		/number=14
FT	exon	3114..3171
FT		/*tag= ae
FT		/number=15
FT	Intron	3172..3247
FT		/*tag= af
FT		/number=15
FT	exon	3248..3321
FT		/*tag= ag
FT		/number=16
FT	Intron	3322..3398
FT		/*tag= ah
FT		/number=16
FT	exon	3399..3423
FT		/*tag= ai
FT		/number=17
FT	Intron	3424..3513
FT		/*tag= aj
FT		/number=17
FT	exon	3514..3700
FT		/*tag= ak
FT		/number=18
FT		3865..3866
FT		/*tag= al
XX		
XX	EP1035206-A1.	
XX		
XX	13-SEP-2000.	
XX		
XX	03-MAR-2000; 2000EP-0104430.	
XX		
XX	09-MAR-1999; 99EP-0104668.	
XX	01-FEB-2000; 2000EP-0101666.	
XX		
XX		
XX	(HOFF) HOFFMANN LA ROCHE & CO AG F.	
XX		
XX	Hoshino T, Ojima K, Setoguchi Y;	
XX		
XX	WPI; 2000-559874/52.	
XX		
XX	P-PSDB; AAB08713.	
XX		
XX		
XX	Novel polynucleotide encoding astaxanthin synthase useful for producing recombinant cells for producing astaxanthin from beta-carotene	
XX		

PS Claim 4; Page 30-33; 46pp; English.

XX The present sequence encodes an astaxanthin synthetase polypeptide

CC of Phaffia rhodozyma. The enzyme is involved in the last step of the

CC astaxanthin biosynthesis pathway, from beta-carotene to astaxanthin.

CC P. rhodozyma is a carotenogenic yeast strain. The astaxanthin

CC synthetase polynucleotides and polypeptides are useful for producing

CC astaxanthin. Astaxanthin is an antioxidant which may be used to

CC protect living cells against diseases such as cancer. Astaxanthin is

CC also used as a colouring reagent, e.g. in farmed fish like salmon to

CC impart an orange-red coloration.

XX

SO Sequence 3969 BP; 937 A; 942 C; 869 G; 1221 T; 0 other;

Query Match 0.6%; Score 18; DB 21; Length 3969;

Best Local Similarity 100.0%; Pred. No. 1.5e+02;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 420 aactgaaagaccctc 437

|||||

Db 1193 AACTTGAAAGCACCTTC 1176

RESULT 73

AA158415 standard; cDNA: 4215 BP.

AA158415;

XX

XX 22-Oct-2001 (first entry)

XX

DE Human polynucleotide SEQ ID NO 618.

XX

KW Human; nootropic; immunosuppressant; cytoskeletal; gene therapy; cancer;

KW peripheral nervous system; neuropathy; central nervous system; CNS;

KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;

KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;

KW chemokine; thrombolytic; drug screening; arthritis; inflammation;

KW leukaemia; ss.

XX

OS Homo sapiens.

XX

PN WO200153312-A1.

XX

PD 26-JUL-2001.

XX

PF 26-DEC-2000; 2000WO-US34263.

XX

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

PR 09-JUL-2000; 2000US-0598042.

PR 19-JUL-2000; 2000US-0620312.

PR 03-AUG-2000; 2000US-0653450.

PR 14-SEP-2000; 2000US-0662191.

PR 19-OCT-2000; 2000US-0693036.

PR 29-NOV-2000; 2000US-0727344.

XX

XX (HYSEQ-) HYSEQ INC.

XX

PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;

PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX

XX WPI: 2001-442253/47.

DR P-PSDB; AAM39259.

XX

PT Novel nucleic acids and polypeptides, useful for treating disorders

PT such as central nervous system injuries -

XX

XX Claim 1; SEQ ID NO 618; 10078pp; English.

XX

CC The invention relates to human nucleic acids (AA157798-AA161363) and

CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,

CC immunosuppressant and cytoskeletal activity. The polynucleotides are useful

CC in gene therapy. A composition containing a polypeptide or polynucleotide

CC of the invention may be used to treat diseases of the peripheral nervous

CC system, such as peripheral nervous injuries, peripheral neuropathy and

CC localised neuropathies and central nervous system diseases, such as

CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic

CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the

CC utilisation of the activities such as: immune system suppression,

CC Activin/inhibin activity, chemotactic/chemokine activity, haemostatic

CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,

CC assays for receptor activity, arthritis and inflammation, leukemias and

CC C.N.S disorders.

CC Note: The sequence data for this patent did not form part of the printed

CC specification.

XX

SO Sequence 4215 BP; 1523 A; 901 C; 855 G; 934 T; 2 other;

Query Match 0.6%; Score 18; DB 22; Length 4215;

Best Local Similarity 100.0%; Pred. No. 1.5e+02;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2831 aaagaacctctgaanaa 2848

|||||

Db 3248 aaagaacctctgaanaa 3265

RESULT 74

AA160201 standard; cDNA: 4871 BP.

AA160201;

XX

XX 22-Oct-2001 (first entry)

XX

DE Human polynucleotide SEQ ID NO 4190.

XX

KW Human; nootropic; immunosuppressant; cytoskeletal; gene therapy; cancer;

KW peripheral nervous system; neuropathy; central nervous system; CNS;

KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;

KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;

KW chemokine; thrombolytic; drug screening; arthritis; inflammation;

KW leukaemia; ss.

XX

OS Homo sapiens.

XX

PN WO200153312-A1.

XX

PD 26-JUL-2001.

XX

PF 26-DEC-2000; 2000WO-US34263.

XX

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

PR 09-JUL-2000; 2000US-0598042.

PR 19-JUL-2000; 2000US-0620312.

PR 03-AUG-2000; 2000US-0653450.

PR 14-SEP-2000; 2000US-0662191.

PR 19-OCT-2000; 2000US-0693036.

PR 29-NOV-2000; 2000US-0727344.

XX

XX (HYSEQ-) HYSEQ INC.

XX

PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;

PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX

XX WPI: 2001-442253/47.

DR P-PSDB; AAM41045.

XX

PT Novel nucleic acids and polypeptides, useful for treating disorders

PT such as central nervous system injuries -

	Pt	Sequence 4871 BP; 1815 A; 1021 C; 990 G; 1045 T; 0 other;	
Cc	Sd	The invention relates to human nucleic acids (AAI57798-AAI61369) and the encoded polypeptides (AAM38642-AAM42213) with nootropic, immunosuppressant and cytostatic activity. The polynucleotides are useful in gene therapy. A composition containing a polypeptide or polynucleotide of the invention may be used to treat diseases of the peripheral nervous system, such as peripheral nervous injuries, peripheral neuropathy and localised neuropathies and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager Syndrome. Other uses include the utilisation of the activities such as: Immune system suppression. Activin/inhibin activity, chemotherapeutic/chemopreventive activity, haemostatic and thrombolytic activity, cancer diagnosis and therapy, drug screening, assays for receptor activity, arthritis and inflammation, leukaemia and C.N.S disorders.	
Cc	Nr	Note: The sequence data for this patent did not form part of the printed specification.	
Xx	Df	Query Match 0.6%; Score 18; DB 22; Length 4871; Best Local Similarity 100.0%; Pred. No. 1.5e+02; Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0	
Oy	Dn	2831 aaagaacctctgaataaa 2848 Db 3258 aaagaccctctgtaaaaa 3275	
Rslult	F5	AAXI3198/C	
Xx	ID	AAXI3198 standard; DNA; 4951 BP.	
Ac	Xx	AAXI3198;	
Xx	Dt	19-MAR-1999 (first entry)	
Xx	De	Enterococcus faecalis genome contig SEQ ID NO:261.	
Kw	Vaccine:	attenuation; computer readable medium; ds.	
Xx	Os	Enterococcus faecalis.	
Xx	Pn	M09850555-A2.	
Xx	Pd	12-NOV-1998.	
Xx	Pf	04-MAY-1998; 98WO-US08985.	
Pr	PR	14-NOV-1997; 97US-0066009.	
Pr	PR	06-MAY-1997; 97US-0044031.	
Pr	PR	16-MAY-1997; 97US-0046655.	
Xx	Pa	(HUMA-) HUMAN GENOME SCI INC.	
Pt	Pt	Barash SC, Dillon PJ, Kunsch CA;	
Xx	WPI:	1999-045171/04.	
Xx	New isolated Enterococcus faecalis polynucleotides and polypeptides		
Pt	- used to develop products for the detection of Enterococcus and for		
Pt	use in vaccines for prevention or attenuation of Enterococcus		
Pt	Infection.		
Xx	Claim 1; Page 1248-1251; 2084pp; English.		
Xx	A computer readable medium has been developed which has recorded on it		
Cc	982 nucleotide sequences isolated from the Enterococcus faecalis genome.		
Cc	AAA12938 to AAA13919 represent these nucleotide sequences which are		

CC primarynucleotide sequences, also known as contigs. The computer-based
CC system can identify fragments of the *Enterococcus faecalis* genome with
CC commercial importance. The products can be used to detect the presence
CC of *Enterococcus faecalis* in samples. They can also be used for
CC diagnosing *Enterococcal* infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of *Enterococcus faecalis*, or
CC another related organism, *in vivo* or *in vitro*. In particular the
CC polypeptides encoded by the *Enterococcus faecalis* nucleotide sequences
CC can be used in vaccines to prevent or attenuate an *Enterococcal*
CC infection.
CC
CC
CC
CC
SQ Sequence 4951 BP; 1563 A; 981 C; 735 G; 1665 T; 7 other;

Query Match 0.6%; Score 18; DB 20; Length 4951;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2527 gctgtctgcgaatgcag 2544
|||||
Db 1288 gcttctctgcgaatgcag 1271

```
Search completed: March 25, 2002, 16:50:28
Job time: 9092 sec
```

THIS PAGE BLANK (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2002, 11:40:31 : Search time 2764.97 Seconds
(without alignments)
12176.090 Million cell updates/sec

Title: US-09-697-089-1

Perfect score: 3133
Sequence: 1 cgcctcgcgcgcggtgggaag.....aagtgctactcgaagccacta 3133

Scoring table:

OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 11351937 seqs, 5372889281 residues

Word size : 0

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estopl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: gp_est1:*
11: gp_est2:*
12: gp_hic:*
13: gp_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pin:*
18: em_gss_pro:*
19: em_gss_rtd:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	551	17.6	552	13	AO309404 CITR1-EL
C 2	446	14.2	480	10	AV719179 AV719179
C 3	364	11.6	364	10	AI263294 Qx57B01.x
C 4	291	9.3	371	10	AV656315 AV656315
C 5	249	7.9	261	10	AW337918 AW337918
C 6	172	5.5	476	11	BG210375 RST29913
C 7	170	5.4	553	13	AO320928 RPTC11-93
C 8	162	5.2	509	10	AW418826 ha21e11.x
C 9	146	4.7	630	13	AO112439 CIT-HSP-2
C 10	124	4.0	412	11	H25984 Y156907.t1
C 11	106	3.4	470	13	AO624020 HS-5378.B
C 12	93	3.0	499	10	AI023795 AO623795

C 13	81	2.6	840	11	BF207840
C 14	58	1.9	219	13	AO283886
C 15	55	1.8	635	10	B58691
C 16	32	1.0	446	10	AI222422
C 17	30	1.0	251	11	BF903662
C 18	28	0.9	404	13	AO889169
C 19	25	0.8	443	13	AO497022
C 20	25	0.8	674	13	AO541295
C 21	23	0.7	606	11	BF829853
C 22	22	0.7	256	10	AV268403
C 23	22	0.7	437	13	AZ740007
C 24	21	0.7	163	13	AZ086396
C 25	21	0.7	343	10	AA784099
C 26	21	0.7	343	10	AA821339
C 27	21	0.7	371	10	A1645009
C 28	21	0.7	432	13	AZ223093
C 29	21	0.7	437	13	AO555443
C 30	21	0.7	462	13	CNS00592
C 31	21	0.7	602	13	AZ360053
C 32	21	0.7	1147	11	BG571228
C 33	20	0.6	64	11	BI142462
C 34	20	0.6	73	10	BE638399
C 35	20	0.6	205	13	AO069062
C 36	20	0.6	215	11	BI315256
C 37	20	0.6	216	10	AA493806
C 38	20	0.6	225	10	A1366637
C 39	20	0.6	248	10	A1603832
C 40	20	0.6	299	13	AZ891870
C 41	20	0.6	332	10	AW357412
C 42	20	0.6	336	13	AO952434
C 43	20	0.6	342	10	AA670222
C 44	20	0.6	380	13	AZ991565
C 45	20	0.6	393	10	AA294641
C 46	20	0.6	393	10	AW838955
C 47	20	0.6	415	10	AA294540
C 48	20	0.6	439	11	T81898
C 49	20	0.6	439	11	AA294143
C 50	20	0.6	483	13	AO209355
C 51	20	0.6	481	11	BF601894
C 52	20	0.6	498	10	A1643174
C 53	20	0.6	505	11	BF602580
C 54	20	0.6	519	11	BG382487
C 55	20	0.6	521	10	A1670486
C 56	20	0.6	521	13	FR0031506
C 57	20	0.6	526	11	BI394563
C 58	20	0.6	535	11	BG351255
C 59	20	0.6	541	13	AZ221490
C 60	20	0.6	554	13	AZ298434
C 61	20	0.6	557	13	AO270247
C 62	20	0.6	571	11	BF258009
C 63	20	0.6	577	11	BG385206
C 64	20	0.6	577	13	AO570315
C 65	20	0.6	580	13	AZ987277
C 66	20	0.6	605	11	BG098273
C 67	20	0.6	620	13	AZ802117
C 68	20	0.6	627	10	A1511575
C 69	20	0.6	633	13	AZ510047
C 70	20	0.6	644	10	AW874886
C 71	20	0.6	652	10	BE346038
C 72	20	0.6	653	10	A1317901
C 73	20	0.6	656	10	BE636495
C 74	20	0.6	660	11	BG310513
C 75	20	0.6	726	13	CNS02935
C 76	20	0.6	729	13	B65395
C 77	20	0.6	731	11	BG250425
C 78	20	0.6	754	13	AO329189
C 79	20	0.6	760	10	A1591096
C 80	20	0.6	772	11	BG310534
C 81	20	0.6	779	11	BG350968
C 82	20	0.6	783	11	BI184833
C 83	20	0.6	784	11	BF824698
C 84	20	0.6	891	11	BG248945
C 85	20	0.6	937	13	CNS02507

BF207840	601862546
AO283886	RPTC11-78
B58691	CIT-HSP-201
AI222422	qh04F06.x
BF903662	IL2-MT018
AO889169	HS-2161.B
AO497022	HS-5197.B
AO541295	RPTC-11-3
BF829853	MR3-HN005
AV268403	AV268403
AZ740007	RPTC-24-1
AZ086396	RPTC-23-2
AA784099	dlg08a1.f
AA821339	vs68a01.f
A1645009	vs68a01.y
AZ223093	RPTC-23-9
AO555443	HS-5224.A
CNS00592	Arabidops
AZ360053	1M0103H1
BG571228	602592168
BI142462	SMOV3MCAM
BE638399	SMOV3MCAM
AO069062	SMOV3MCAM
BI315256	da96e06.f
AA493806	h02f04.s
A1366637	SMOV3MCAM
A1603832	SMOV3MCAM
AZ891870	RPTC-24-1
AW357412	40862.MAR
AO952434	Sheared.D
AA670222	ad19h11.s
AZ991565	2M0275B24
AA294641	SMOV3MCAM
AW838955	CM2-LT006
AA294540	SMOV3MCAM
T81898	Yd29a11.s1
AA294143	SMOV3MCAM
AO209355	HS-3740.A
BF601894	266990.MA
A1643174	v946g09.y
BF602580	267845.MA
BG382487	298398.MA
A1670486	SMOV3MCAM
AI027875	Fugu.rubr
BI394563	pap1n.pko
BG351255	088D06.Ma
AZ221490	GM-UMB001
AZ298434	RPTC-23-1
AO270247	HS-2045.A
BF258009	HYSMET001
BG385206	306845.MA
AO570315	HS-5352.B
AZ987277	EST4629J15
BG098273	2M02629J15
AZ802117	2M0061H04
A1511575	SMOV3MCAM
AZ510047	1M0354H14
AW874886	SMOV3MCAM
BE346038	SMOV3MCAM
A1317901	SMOV3MCAM
BE636495	SMOV3MCAM
BG310513	SMOV3MCAM
AL186746	Tetradodon
B65395	CIT-HSP-202
BG250425	602362580
AO329189	nbx00044F
A1591096	tbw1B09.x
BG310534	SMOV3MCAM
BG350968	099C11.MA
BI184833	UNL-P-FN-
BF824698	SMOV3MCAM
BG248945	602361385
AL181456	Tetradodon

/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2: XhoI"
 BASE COUNT 138 a 120 c 85 g 137 t
 ORIGIN

Query Match 14.2%; Score 446; DB 10; Length 480;
 Best Local Similarity 100.0%; Pred. No. 7.4e-214;
 Matches 446; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2640 catgaactgacagacagatgaacgtgctagaaacagctaccgactgactgctccctg 2699
 |||||||
 DB 480 CATGAACCTGATCGACAGATGAACGCTAGAAACACCTCACCTGATGCTGCCCTGG 421
 |||||||
 QY 2700 gactggaagctgcaagcagcctgagcaacccgtgtgaacatttggaagaggtcccaaa 2759
 |||||||
 DB 420 GGGTGGAGCTGCAAGCGACCTGAGACCTGTTGAACATTGGAGAGGTCACCA 361
 |||||||
 QY 2760 cctgcaagcttgggttgtaaaaactgagacacacagatagagattgaattagtt 2819
 |||||||
 DB 360 CTCGTCAACCTTGGCTTGAACAACTGAGACCTCACAGATACAGATTAGATTAGT 301
 |||||||
 QY 2820 gcaatttttggaagaaccctctgaaaaactccagcagltgaaatttggcgggaaatcgt 2879
 |||||||
 DB 300 GCATTTTGGAAAGAACCCCTCGAATAAATTCACAGATTGAATTTGGCGGGAATCGT 241
 |||||||
 QY 2880 gtgagcagtgatgagtgcttgccttcacatggtgtgtattgagaatttaagcattagtt 2939
 |||||||
 DB 240 GTGAGCAGTGAATGATGCTGCTTCATGAGGGTGTATTGAGAACTTTAAGCAATTAGTG 181
 |||||||
 QY 2940 tttttgacttagtactaaagaatttctacactgacacagcattagtcagaaactagc 2999
 |||||||
 DB 180 TTTTGGACTTTAGTACTTAAGAAATTTCTACCGATCCAGCATTAAGCAAAAACCTAGC 121
 |||||||
 QY 3000 caagtgtatccaagtttaacttcttcgcaagaagcagctgctgtggtggcaattgcat 3059
 |||||||
 DB 120 CAAGTGTATTCCAAGTTAATCTTTTCGCAAGAGCTAGGCTTGTGGTGGCAATTTGAT 61
 |||||||
 QY 3060 gatgatgactcagtgatttactagc 3085
 |||||||
 DB 60 GATGATGATCTCAGTGTATTACAG 35

RESULT 3
 LOCUS A1263294 364 bp mRNA EST 03-FEB-1999
 DEFINITION qx57b01.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2005417 3',
 mRNA sequence.
 ACCESSION A1263294
 VERSION A1263294.1 GI:3871497
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 364)
 REFERENCE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap
 TITLES National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaabs-r@mail.nih.gov
 Life Technologies catalog #: 11548-013
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/BLN at:
 www.bio.lnl.gov/bbrp/image/image.html
 Insert Length: 2146 Std Error: 0.00
 Seq primer: -40UP from Gibco
 High quality sequence stop: 364.
 Location/Qualifiers
 1. 364
 /organism="Homo sapiens"

/db_xref="taxon:9606"
 /clone_image="2005417"
 /clone_lib="NCI_CGAP_Pan1"
 /tissue_type="adenocarcinoma"
 /lab_host="DH10B"
 /note="Organ: pancreas; Vector: PCMV-SPORE6; Site_1: SalI;
 Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
 Average insert size 1.72 kb. Life Technologies catalog #:
 11548-013"
 BASE COUNT 117 a 84 c 55 g 108 t
 ORIGIN

Query Match 11.6%; Score 364; DB 10; Length 364;
 Best Local Similarity 100.0%; Pred. No. 2e-172;
 Matches 364; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2766 aagcttgggttgtaaaaactgagacacacagatagagaattgaattagtgcaatt 2825
 |||||||
 DB 364 AAGCTTGGGTTGAACAACTGAGACCTCACAGATACAGATTAGATTAGTGCATT 305
 |||||||
 QY 2826 ttggaagaagaaccctctgaaaaactccagcagltgaaatttggcgggaaatcgtgtgagc 2885
 |||||||
 DB 304 TTTGGAAGAAAGAACCCCTCGAATAAATTCACAGATTGAATTTGGCGGGAATCGTGTGAGC 245
 |||||||
 QY 2886 agtgaatgagatgcttgccttcacatggtgtgtattgagaatttaagcattagttttt 2945
 |||||||
 DB 244 AGTGATGATGATGCTTGCCTTCATGAGGGTGTATTGAGAACTTTAAGCAATTAGTGT 185
 |||||||
 QY 2946 gacttagtactaaagaatttctacactgacacagcattagtcagaaacttagccaaagt 3005
 |||||||
 DB 184 GACTTAGTACTTAAGAAATTTCTACCGATCCAGCATTAAGCAAAAACCTAGCAAGTG 125
 |||||||
 QY 3006 ttatccaagttaacttcttcgcaagaagcagctgtgtggtggcaatttgaatgatgat 3065
 |||||||
 DB 124 TTATCCAAAGTTAATCTTTTCGCAAGAGCTAGGCTTGTGGTGGCAATTTGATGATGAT 65
 |||||||
 QY 3066 gatcagtgatttactagtgctttaaactagtaactgcttaataaagtgtactcga 3125
 |||||||
 DB 64 GATCTCAGTGTATTACAGGTGCTTTTAACTAGTAATCTTAATTAAGGTACTCGA 5
 |||||||
 QY 3126 agcc 3129
 |||||
 DB 4 AGCC 1

RESULT 4
 LOCUS AV656315 371 bp mRNA EST 07-SEP-2000
 DEFINITION AV656315 GUC Homo sapiens cDNA clone G1CEGA10 3', mRNA sequence.
 ACCESSION AV656315
 VERSION AV656315.1 GI:9877329
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 371)
 REFERENCE Qian,B., Wu,T., Huang,Q., Huang,C., Kang,B., Gao,X., Xu,Z., Xiao,H.,
 Xu,X., Li,N., Peng,Y., Liu,F., Qu,J., Song,H., Cheng,Z., Qu,J.,
 Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu
 G., Yang,Y., Gu,Y., Chen,Z. and Han,Z.
 Homo sapiens cDNA clone
 JOURNAL Unpublished (2000)
 COMMENT Contact: Zeguang Han
 Chinese National Human Genome Center at Shanghai
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
 201203, P. R. China
 Tel: 86-21-50801919(ex.45)
 Fax: 86-21-50801922
 Email: hanzg@cnhg.sh.cn
 This clone is available at CHGC in Shanghai.
 Location/Qualifiers

FEATURES
 source 1. 364
 /organism="Homo sapiens"

source 1..371
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GLCEQA10"
/clone_lib="GLC"
/tissue_type="corresponding non cancerous liver tissue"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pbluescript sk(-); Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 112 a 85 c 91 g 82 t 1 others
ORIGIN

Query Match 9.3%; Score 291; DB 10; Length 371;
Best Local Similarity 99.7%; Pred. No. 1.5e-135;
Matches 341; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1908 gaaagagctgagaagaacacagggggaatccacatggaagagcccaagaacacatt 1967
|||||
Db 1 GAAAGGCTGAGAGAGACAGGTGGATCCACATGGAAGGCCCCAGAAACTACAT 60
|||||

Qy 1968 cccagcagggctgatactcttctcactggaagcaggaattcaggaactctggaagtc 2027
|||||

Db 61 CCCAGCAGGCTGATCTTTGTTCTTCACTGGAAGCAGAACTCAGACTCTGAGGTC 120
|||||

Qy 2028 acactccggagattcagaagttgataagaagatacacatcctgagggaataatc 2087
|||||

Db 121 ACACCTCCGGATTCAGCAAGTTGAATAGCAAGATACAGATATCTGGGAAAAATATTC 180
|||||

Qy 2088 agctctgcacaaagcctcagagctgcaaatgaagagatgctggtgctggaagcctc 2147
|||||

Db 181 AGCTCTGCCAAGAGCTCAGGCTGCAGAAATGAAGATGTGCTGGTGGAGAGCCTC 240
|||||

Qy 2148 agttgtccctcagcagcctgaaagacattatctcctcagtggtggaagcagtcctc 2207
|||||

Db 241 AGTTGGCTCCTCAGCAGCTGTGAAGACATTTATCTCTCATGTGGAGGCCAGTCCCTC 300
|||||

Qy 2208 accatagaagatgagagacacatcacatctgtaacaaactg 2249
|||||

Db 301 ACCATAGAAAGATGAGAGCAGCATCATCTGTAAACAAACCTG 342
|||||

RESULT 5
AW337918/c 261 bp mRNA EST 31-JAN-2000
LOCUS he12h11.x1 NCI_CGAP_CML1 Homo sapiens cDNA clone IMAGE:2918853 3'
DEFINITION mRNA sequence.
ACCESSION AW337918
VERSION AW337918.1 GI:6834544
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 261)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-rt@mail.nih.gov
Tissue Procurement: Elisabeth Paletta, Jonathan D. Licht, M.D.,
Michael R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: Life
Technologies, Inc. CDNA Library Arrayed by: Christa Prange, The
I.M.A.G.E. Consortium DNA Sequencing by: Washington University
Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www-bio.llnl.gov/bdrrp/image/image.html
Seq primer: -40up from Gibco
High quality sequence stop: 201.
Location/Qualifiers

FEATURES
1..476
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"

source 1..261
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2918853"
/clone_lib="NCI_CGAP_CML1"
/tissue_type="myeloid cells, 18 pooled CML cases, BCR/ABL
rearrangement positive, includes both chronic phase and
myeloid blast crisis"
/lab_host="DH10B"
/note="Organ: whole blood; Vector: pCMV-SPORT6; Site_1:
SalI; Site_2: NotI; Cloned unidirectionally. Primer:
Oligo dT. Library constructed by Life Technologies."

BASE COUNT 86 a 54 c 38 g 83 t
ORIGIN

Query Match 7.9%; Score 249; DB 10; Length 261;
Best Local Similarity 100.0%; Pred. No. 2.3e-114;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2885 cagtatgagatgctgcttcctcatggtgattgagaatcttaagcaattagtgttt 2944
|||||

Db 261 CAGTATGAGATGCTGCTCCCTTCATGCGTATTTAGAAATCTTAAGCAATTAGTGTTTT 202
|||||

Qy 2945 tgactttagtaagaattcttaccctgataccagcattagtcagaaactagccaagt 3004
|||||

Db 201 TGACTTGTAGTAAAGAAATTTCTACCTGATCCAGATAGTCAGAAAACCTTAGCCAGT 142
|||||

Qy 3005 gtatccaaagttaactttctgcagaagaagcagctgctgttggtgccaatttgatgaga 3064
|||||

Db 141 GTATCCAAAGTTAACTTTCTGCAAGAGCTAGGCTTGTGGCTGCAATTTGAGATGAGA 82
|||||

Qy 3065 tgatccagtggttattcaagtgcttcttaactagtaactgcttaataagtgctacg 3124
|||||

Db 81 TGATCTGATGTTATTATTAAGAGTGTCTTTTAACCTAGTAACCTCTTAATTAAGTACTCG 22
|||||

Qy 3125 aagccagta 3133
|||||

Db 21 AAGCCAGTA 13
|||||

RESULT 6
BG210375 476 bp mRNA EST 21-APR-2001
LOCUS RST29913 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
DEFINITION BG210375
ACCESSION BG210375
VERSION BG210375.1 GI:13732062
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 476)
Harrington, J.J., Sherf, B., Rundlett, S., Jackson, P.D., Perry, R.,
Cain, S., Dahl, T., Thornton, M., Ramachandran, R., Whittington, J.,
Lerner, L., Krasnoc, D., McElligott, K., Clark, S., Mays, R., Smith, E.,
Velasco, N., Hess, J., Cothren, K., Lo, K., Offenbacher, J., Danzig, J.,
and Ducar, M.
JOURNAL Creation of Genome-wide Protein Expression Libraries using Random
Activation of Gene Expression
Nat. Biotechnol. 19 (5), 440 (2001) In press
COMMENT Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scain@athersys.com
High quality sequence stop: 360.
Location/Qualifiers

FEATURES
1..476
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"

	Query Match	Best Local Similarity	5.5%: Score 172; DB 11;	Length 476;
Db	242	TGGGCTATTTCAGAAATCTTAAGCAATTAAGTCTTTTTCACCTTTAGTACTAAAGATTTC	99.6%: Pred. No. 2.1e-75;	
		Matches 222; Conservative 0; Mismatches 1; Indels 0; Gaps 0;		
Qy	2908	tggtgtatcttgagaatcttaagcaatcagtcgtcttttgaacttgaacttaagaatctc		2967
Db	242	TGGGCTATTTCAGAAATCTTAAGCAATTAAGTCTTTTTCACCTTTAGTACTAAAGATTTC		301
Qy	2968	tacctgatccagcatttagtcagaaaaacttagccaaagtgtlctlccaaagttaactttctgc		3027
Db	302	TACCGATTCACGATTTAGTCAGAAAACTTAGCCAAAGTGTTATCCAAAGTTAATCTTCTGC		361
Qy	3028	aagaagctcagagctctgtgtgggtggcgaatttgatgatgatgatctccagctgttalttacaagtgtg		3087
Db	362	AAGAAGCTTAGGCTTGTGTGGGTGGCAATTGTGATGATGATGATCTCAGTGCCTATTACAGGTG		421
Qy	3088	ctttaaactagtaactgcttaaaaaaagtgtaactcgaagcca		3130
Db	422	CTTTTAAACTAGTAACTGCTTTAAATTAAGTGATGACGGAAGCCA		464

RESULT	7
LOCUS	AQ320928/c
DEFINITION	AQ320928 553 bp DNA GSS 06-MAY-1999 RPC111-93C9.TV RPC1-11 Homo sapiens genomic clone RPC1-11-93C9, DNMA
ACCESSION	AQ320928
VERSION	AQ320928
KEYWORDS	GSS.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 553)
AUTHORS	Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., White,C., de Jong,P. and Venter,D.C. Use of human BAC End Sequences for Sequence-Ready Map Building Unpublished (1998) Other-GSSs: RPC111-93C9.TV Contact: Shaying Zhao, William Niernan, Mark Adams Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850 Tel.: 301 838 0200 Fax: 301 838 0208
TITLE	Email: hbe@tigr.org
JOURNAL	Clones are derived from the human BAC library RPC1-11. For BAC library availability, please contact Pieter de Jong (pietere@jeng.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html Seq primer: T7
COMMENT	Class: BAC ends. Location/Qualifiers 1..553 /organism="Homo sapiens" /db_xref="GDB:7535384" /db_xref="taxon:9606" /clone="RPC1-11-93C9" /clone_1lb="RPC1-11" /sex="Male"

	Query Match	Best Local Similarity	Matches 170: Conservative	5 4%: Score 170: DB 13: Length 553;	100.0%: Pred. NO. 2.1e-74;	0: Mismatches 0; Indels 0; Gaps 0;
OY	2650	tcgaacagatgaacgtgcgtactagaacagcctccacgcgaactgctgcgccctggagcttgacg	170 a	107 c	114 g	162 t
Db	258	TCGACAGGATGAGACTGCTGTAAGAACAGCTCACCGCACACTATGCTGCCCTGGGCGCTGTGACG				
OY	2710	tgcaagcagcgcctgagcagcctgtgtgaacaatttggagagagtlccacaaactcgtlcaagc				
Db	198	TGCAAGCGCAGCCTGGAGCAGCGTGTGAAACAATTGTGGAGAGAGTCCCCACAACTGCTCAACG				
OY	2770	ttgggttgaanaactgtagagactcacagaatacagaattagaaatttagt				
Db	138	TTGGGTTGAAACTGAGACTCACAGATACAGAGATTGAAATTTTAGT				

LOCUS	509 bp	mRNA	EST	09-FEB-2000
ACCESSION	AW18826	haz1e11.x1	NCI_CGAP_K1d12	Homo sapiens cDNA clone IMAGE:2874380 3'
VERSION	AW18826			
KEYWORDS		mRNA sequence.		
SOURCE	AW18826.1	GI:6946758		
ORGANISM	human.			
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
COMMENT				
FEATURES				
SOURCE				

Db	124	TGGTGTGGCTGGAAGCCTCAGCTTTGGTCTCAGCAGCCTGTAAAGAAC	169
RESULT	12		
LOCUS	A1023795	499 bp	EST
DEFINITION	ox08d03.x1 Soares_fetal_liver_spleen_INFLS_S1 Homo sapiens CDNA clone IMAGE:1655717 3', mRNA sequence.		28-AUG-1998
VERSION	A1023795		
KEYWORDS	A1023795.1 GI:3238839		
SOURCE	EST.		
ORGANISM	human.		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 499)		
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .		
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index		
COMMENT	Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Insert Length: 1179 Std. Error: 0.00 Seq primer: -40m13 fwd. ET from Amersham High quality sequence stop: 439. Location/Qualifiers 1. 499 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:1655717" /clone_id="Soares_fetal_liver_spleen_INFLS_S1" /sex="male" /dev_stage="20 week-post conception fetus" /lab_host="DH10B (ampicillin resistant)" /note="Organ: Liver and Spleen; Vector: p773D (Pharmacia) with a modified polylinker; Site_1: Pac I; Site_2: Eco RI; This is a subtracted version of the original Soares fetal liver spleen INFLS library. 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5', AACGGAAGATTAATTAAGACCTCTTTTCTTTTCTTTTCTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified p773 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."		
BASE COUNT	144 a 109 c 98 g 148 t		
ORIGIN			
Query Match	3.0%; Score 93; DB 10; Length 499;		
Best Local Similarity	100.0%; Pred. No. 1.7e-35;		
Matches	93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	298 gctcttccatcagacacagaaagagacttgacagattggctcaggaattaaagact 357		
Db	407 GTCCTTTTCATCAGACATCAGAGAGAGACTTGGCAGATTGAGATTAAAGGACT 466		
QY	358 tttaccatacccatcttctgaactttatc 390		
Db	467 TGTACCATACCCCATCTTTTGTGAACTTTATC 499		
RESULT	13		
LOCUS	BF207840/c	840 bp	mRNA
DEFINITION	601865546F1 NIH_MGC_53 Homo sapiens CDNA clone IMAGE:4082074 5', mRNA sequence.		EST
ACCESSION	BF207840		
VERSION	BF207840.1	GI:11101426	
KEYWORDS	EST.		
SOURCE	human.		

ORGANISM	Homio sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE	1 (bases 1 to 840)
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/ .
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strusberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: CLONTECH Laboratories, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov Plate: L10M944 row: 1 column: 11 High quality sequence stop: 656. Location/Qualifiers
FEATURES	1..840
SOURCE	<p>/organism="Homo sapiens"</p> <p>/db_xref="taxon:9606"</p> <p>/clone="IMAGE:4082074"</p> <p>/clone_lib="NIH-MGC_53"</p> <p>/tissue_type="carcinoma, cell line"</p> <p>/lab_host="DH10B (T1 phage-resistant)"</p> <p>/note="Organ: bladder; Vector: pDNR-LIB (Clontech); Site_1: SfiI (ggcgctcgcc); Site_2: SfiI (ggccatattggc); Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGCGC-3' and 3' adaptor sequence: 5'-ATCTAGAGCGCGAGCGCGCATCATG-drr(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.55 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."</p>
BASE COUNT	276 a 143 c 165 g 256 t
ORIGIN	
Query Match	2.6%; Score 81; DB 11; Length 840;
Best Local Similarity	100.0%; Pred. No. 2e-29;
Matches: 81; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
OY 3053	attgagatgatgattcctcagtggtattacagggctttaactagtaactgcttaaat 3112
Db 518	ATTGATGATGATGATGATTCACAGTGTTATTCACAGTGCTTTAACTAGTAAGCTTAAT 459
OY 3113	aaagtgtaactggaagccagta 3133
Db 458	AAAGTGTAAGCTGGAAGCCAGTA 438
SULT 14	
LOCUS 283886	
DEFINITION	AQ283886 219 bp DNA GSS 27-APR-1999 RP111-78E13, TV RPT-11 Homo sapiens genomic clone RPT-11-78E13, DNA sequence.
ACCESSION	AQ283886
VERSION	AQ283886.1 GI:3910204
KEYWORDS	GSS.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE	1 (bases 1 to 219)
JOURNAL	Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,U.C. Use of human BAC End Sequences for Sequence-Ready Map Building Unpublished (1998)
COMMENT	Contact: Mark Adams Department of Eukaryotic Genomics The Institute for Genomic Research

```

9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@edlong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: 17
Class: BAC ends.

FEATURES
Source
Location/Qualifiers
1..219
/organism="Homo sapiens"
/db_xref="GDB:7529676"
/db_xref="taxon:9606"
/clone="RPCI-11-78E13"
/clone_id="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/vector="pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
RPC11 Human Male BAC library"
BASE COUNT 69 a 45 c 46 g 59 t
ORIGIN
Query Match 1.9%; Score 58; DB 13; Length 219;
Best Local Similarity 100.0%; Pred. No. 8.1e-18;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY 2237 tgtatacaaccctggaaccttgatgattatcattacacgaatcaacgctgccggt 2294
|||||
Db 104 tttatacaaacctggaaaccttgatgattatcattacacgaatcaacgctgccggt 161

RESULT 15
LOCUS B58691 635 bp DNA GSS 20-JUN-1998
DEFINITION CIT-HSP-2013L21.TF CIT-HSP Homo sapiens genomic clone 2013L21, DNA
sequence.
ACCESSION B58691
VERSION B58691.1 GI:2613409
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 635)
Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden
,K., Berry,K., Granger,D., Sun,E., Wible,C., Shizuya,H., Simon,M.
and Venter,J.C.
Use of a random BAC End Sequence Database for Sequence-Ready Map
Building
Unpublished (1997)
Other_GSS: CIT-HSP-2013L21.TF
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: M13-21
Class: BAC ends.
Location/Qualifiers
1..635
/organism="Homo sapiens"
/db_xref="GDB:7042463"
/db_xref="taxon:9606"
FEATURES
Source

```

BASE COUNT	ORIGIN	217	a	122	c	90	g	206	t
		<pre> /clone_1ib="CIT-HSP" /sex="Male" /cell_type="Sperm" /ote="Vector: pBelobAC11; site_1: HindIII; site_2: HindIII" </pre>							

Query Match	1.8%	Score 55:	DB 13:	length 635:					
Best Local Similarity	100.0%	Pred. No.	2.7e-16:						
Matches	55:	Conservative	0:	Mismatches	0:	Indels	0:	Gaps	0:

QY 2765 caagcttggttgaaaaactgagactcacagatacagagattgaatttagt 281g

Db 1 CAAGCTTGCGTTGAAAAACTGGAGACTCACAGATACAGAGAGATTGAATTTTAGGT 55

RESULT	16
LOCUS	AI222422
DEFINITION	AI222422 446 bp mRNA EST 30-NOV-1998 gi004f06.x1 Soares_NFL_T-SHC_S1 Homo sapiens cDNA clone IMAGE:1843715 3', mRNA sequence.
ACCESSION	AI222422
VERSION	AI222422.1 GI:3804625
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi
Mammalia: Eutheria: Primates: Catarrhini, Homnidae: Homo.
1 (bases 1 to 446)
NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lln.gov) for further information.
Insert Length: 546 Std Error: 0.00
Seq primer: -40UP from Glbco
High quality sequence stop: 437.

FEATURES

source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone IMAGE:1843715"
/clone_11d="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7AD-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco R1;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHT, and B-cell
NCL-GCAP-GCBI) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731359. Subtraction by Bento
Soares and M. Fatima Bonaldi."

```

Query Match	1.0%	Score 32:	DB 10:	Length 446;
Best Local Similarity	100.0%	Pred. No. 0.00011;		
Matches	32:	Conservative	0;	Mismatches 0;
			Indels	0;
			Gaps	0
QY	236	gctcttttcacagacatcaagaagagactg	329	
Db	415	gctcttttcacagacatcaagaagagactg	446	

RESULT	17
LOCUS	BF903662
DEFINITION	BF903662 251 bp mRNA
ACCESSION	U12-MT0180-181200-276-F03 MT0180 Homo sapiens
VERSION	BF903662
KEYWORDS	BF903662.1 GI:12295121
SOURCE	EST.
	human.

REFERENCE
1 (bases 1 to 251)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

AUTHORS	TITLE
Dias Neto, E., Garcia Correa, R., Vertovski-Almeida, S., Bittencourt, M.R., Negal, M.A., de Silva, W. Jr., Zago, M.A., Bordim, S., Costa, F. F., Goldman, G.H., Carvalho, A.F., Mutsaers, A., Bala, G.S., Simpson, D.H., Bruns, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL MEDLINE	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
COMMENT	20202663 Contact: Simpson A.J.G.

Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil

Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?pl1=ll2&t2=ll2-MT0180>)
181200-276-F03&t3=2000-12-18&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 45
High quality sequence stop: 96.

FEATURES
source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MT0180"
/dev_stage="Adult"
/note="Organ: marrow; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESFES PCR (U.S. Letters Patent application No. 1,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT
ORIGIN
72 a      62 c      56 g      61 t

```

Query	March	1.0%;	Score	30;	DB	11;	Length	251;	
	Best Local Similarity	100.0%;	Pred. No.	0	0.0011;				
Matches	30;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0
QY	1092	gagttcacactctcacacaacaacagctg	1121						
Db	52	GAGTTCACCTCTCACACACAAACCAACCTG	81						

RESULT	18
AO889169/c	
LOCUS	404 bp DNA
DEFINITION	GSS
ACCESSION	HS-2161.B1.A01.T7C CTR Approved Human Genomic Sperm Library D
VERSION	AO889169 sapiens genomic clone Plate=2161 Col=1 Row=B, DNA sequence.
KEYWORDS	AO889169.1 GI:6345359
	GSS.

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T., Kellier, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and Hood, L.
AUTHORS

TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: <http://www.htsc.washington.edu>
Plate: 2161 row: B column: 1
Seq primer: T7
Class: BAC ends
High quality sequence stop: 404.

FEATURES
SOURCE Location/Qualifiers
1..404
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2161 Col=1 Row=B"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelBAC11; BAC Clones in E-Coli DH10B"
BASE COUNT 119 a 93 c 70 g 122 t
ORIGIN

Query Match 0.9%; Score 28; DB 13; Length 404;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 188 tagagggatcattcagatgatttga 215
|||||
Db 200 TAGAGGATCATTCACTGATTTGAAA 173

RESULT 19
AO497022/c 443 bp DNA GSS 28-APR-1999
LOCUS HS_5197_B1_C11.SP6E RPCI-11 Human Male BAC Library Homo sapiens
DEFINITION AO497022
ACCESSION AO497022
VERSION AO497022.1 GI:4697145
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T., Kellier, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and Hood, L.
AUTHORS

TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887

EMAIL: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Research Genetics (info@resgen.com). BAC end Web Server: <http://www.htsc.washington.edu>
Plate: 773 row: F column: 21
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 443.

FEATURES
SOURCE Location/Qualifiers
1..443
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=773 Col=21 Row=F"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI methylase. Size selected DNA was cloned into the pBACe3.6 vector at EcoRI sites"
BASE COUNT 148 a 90 c 75 g 129 t 1 others
ORIGIN

Query Match 0.8%; Score 25; DB 13; Length 443;
Best Local Similarity 100.0%; Pred. No. 0.39;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 12 ggtgggaagcttcacgcagaa 36
|||||
Db 180 GGTGGGAAGCTTCACTCAGAACAA 156

RESULT 20
AO541295/c 674 bp DNA GSS 19-MAY-1999
LOCUS AO541295
DEFINITION RPCI-11-343A21.TV RPCI-11 Homo sapiens genomic clone RPCI-11-343A21
ACCESSION AO541295
VERSION AO541295.1 GI:4871751
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T., Kellier, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D., Nierman, W., Malek, J., de Jong, P. and Venter, J.C.
AUTHORS

TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building
JOURNAL Unpublished (1997)
COMMENT Other-GSS: RPCI-11-343A21.TV
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers
1..674
/organism="Homo sapiens"


```

/db_xref="GDB:7631348"
/db_xref="taxon:9606"
/clone_1lb="RPCL-11-343A21"
/clone_1lb="RPCL-11"
/sex="Male"
/cell_type="Lymphocytes"
/Note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
RPC11 Human Male BAC Library"
BASE COUNT      218 a      138 c      133 g      185 t
ORIGIN

```

```

Query Match      0.8%; Score 25; DB 13; Length 674;
Best Local Similarity 100.0%; Pred. No. 0.39;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

12 ggtggaagcttcacagacaa 36
|||||
Db      170 GGTGGAGCTTCATCCAGACAA 146

RESULT 21
BF829853      606 bp      mRNA      EST      13-JAN-2001
LOCUS      MR3-HN0052-261200-002-cl2 HN0052 Homo sapiens cDNA, mRNA sequence.
DEFINITION      BF829853
ACCESSION      BF829853
VERSION      BF829853.1 GI:12175855
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 606)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?rl=MR3&L2=MR3-HN0052-
261200-002-cl2&L3=2000-12-26&L4=1)
Seq primer: puc 18 forward
High quality sequence start: 24
High quality sequence stop: 538.
Location/Qualifiers
1. 606
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1lb="HN0052"
/dev_stage="Adult"
/Note="Organ: head,normal; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

```

```

BASE COUNT      119 a      177 c      195 g      113 t      2 others
ORIGIN

```

```

Query Match      0.7%; Score 23; DB 11; Length 606;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

1922 agaacagcgtgcacatcg 1944
|||||
Db      100 AGACACAGGTGCATCCACATGG 122

RESULT 22
AV268403      256 bp      mRNA      EST      05-NOV-1999
LOCUS      AV268403 RIKEN full-length enriched, adult male testis (DH10B) Mus
DEFINITION      AV268403
ACCESSION      AV268403
VERSION      AV268403.1 GI:6256440
KEYWORDS      EST.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 256)
Kono,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai
,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusabe,M.,
Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata
,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H.,
Suzuki,H., Takahashi,F., Tateo,M., Tomianga,N., Tsunoda,Y.,
Wataniki,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T.,
Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Kono,H., et al. 1999)
Unpublished (1999)
Contact: Yoshinide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Sasaki,N., Izawa,M., Wataniki,M., Ozawa,K., Tanaka,T., Yoneda,Y.,
Matsura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and Hayashizaki
,Y.
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
Itoh,M., Kitsuai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.
Location/Qualifiers
1. 256
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_1lb="4930534P12"
/clone_1lb="RIKEN full-length enriched, adult male testis
(DH10B)"
/sex="male"
/tissue_type="testis"
/dev_stage="adult"
/lab_host="DH10B"
/Note="Site_1: SalI; Site_2: BamHI; cDNA library was

```

prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5',
GAGGAGAGAGAGATCCACAGCTCTTTTCTTTTCTTTTNN 3'], cDNA was prepared by using Trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5',
GAGGAGAGATTCGAGTATTAATTAATCCGCCGCCGCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified plasmid pUC(+/-) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI."

BASE COUNT 80 a 36 c 31 g 109 t

ORIGIN

Query Match 0.7%; Score 22; DB 10; Length 256;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2590 ttcttgattatcagaataa 2611
|||||
45 TTCTTGATTATTCAGAAATTA 66

RESULT 23
LOCUS A2740007 437 bp DNA GSS 25-JAN-2001
DEFINITION RPCI-24-103A13.TV RPCI-24 Mus musculus genomic clone RPCI-24-103A13
, DNA sequence.
ACCESSION A2740007
VERSION A2740007.1 GI:12510653
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. 1 (bases 1 to 437)
AUTHORS Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akintet, B., Levins, M., Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregorjais, E., Russell, D., de Jong, P. and Fraser, C.M.
TITLE Mouse BAC End Sequences from Library RPCI-24
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: RPCI-24-103A13.TVB
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhaoc@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cno.org). Clones may be purchased from BACPAC Resources (<http://www.chori.org/bacpac/orderingframe.htm>). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 103 row: A column: 13
Seq primer: T7
Class: BAC ends.

FEATURES
source Location/Qualifiers
1. 437
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-103A13"
/clone_1id="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pFARBAC1; Site_1: BamHI; Site_2: BamHI; RPCI-24 Mouse BAC Library produced by Pieter de Jong. The

library was cloned in the pFARBAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."

BASE COUNT 194 a 81 c 63 g 99 t

ORIGIN

Query Match 0.7%; Score 22; DB 13; Length 437;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1151 gaaacaacacacacataaa 1172
|||||
Db 204 GAAACAACAAACAAACATMAA 225

RESULT 24
LOCUS A2086396 163 bp DNA GSS 08-MAY-2000
DEFINITION RPCI-23-26D1.TV RPCI-23 Mus musculus genomic clone RPCI-23-26D1, DNA sequence.
ACCESSION A2086396
VERSION A2086396.1 GI:7728130
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 163)
AUTHORS Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akintet, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P. and Fraser, C.M.
TITLE Mouse BAC End Sequences from Library RPCI-23
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: RPCI-23-26D1.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhaoc@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics (info@resen.com). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 26 row: D column: 1
Seq primer: T7
Class: BAC ends.

FEATURES
source Location/Qualifiers
1. 163
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-26D1"
/clone_1id="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methyase." Site selected DNA was cloned into the pBAC3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 57 a 32 c 21 g 53 t

ORIGIN

Query Match 0.7%; Score 21; DB 13; Length 163;
Best Local Similarity 100.0%; Pred. No. 40;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1152 aaaaacaacacacacataaa 1172
 Db 140 AAAACAACACACACACATAAA 160

RESULT 25
 AA784099 343 bp mRNA EST 29-JUL-1998
 LOCUS AA784099
 DEFINITION d1908a1.f1 Aspergillus nidulans 24hr asexual developmental and vegetative cDNA lambda zap library Emericella nidulans cDNA clone d1908a1 3', mRNA sequence.

ACCESSION AA784099
 VERSION AA784099.1 GI:2844267
 KEYWORDS EST.
 SOURCE Emericella nidulans.
 ORGANISM Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Emericella; Trichocomaceae; Emericella.

REFERENCE 1 (bases 1 to 343)
 Kupfer,D., Gray,D., Hausner,J., Lai,H., Martin,W., Aramayo,R., Prade,R. and Roe,B.
 An Aspergillus nidulans EST Database
 Unpublished (1998)
 Other ESTs: d1908a1.r1
 Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
 Department of Chemistry and Biochemistry
 Advanced Center for Genome Technology, University of Oklahoma
 620 Parrington Oval, Norman, OK 73019, USA
 Tel: 405 325 4912
 Fax: 405 325 7762
 Email: broe@ou.edu
 We anticipate the future release of the cDNA clones to the Fungal Genetics Stock Center
 Seq primer: M13-20
 High quality sequence stop: 301.
 Location/Qualifiers
 1..343
 /organism="Emericella nidulans"
 /strain="FGSC A26"
 /db_xref="taxon:162425"
 /clone="d1908a1"
 /clone_lib="Aspergillus nidulans 24hr asexual developmental and vegetative cDNA lambda zap library"
 /tissue_type="vegetative mycelia, asexual structures"
 /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript 3' end of cDNA cloned into XhoI site of pBluescript"

BASE COUNT 85 a 94 c 75 g 89 t

ORIGIN
 source
 1..343
 /organism="Emericella nidulans"
 /strain="FGSC A26"
 /db_xref="taxon:162425"
 /clone="d1908a1"
 /clone_lib="Aspergillus nidulans 24hr asexual developmental and vegetative cDNA lambda zap library"
 /tissue_type="vegetative mycelia, asexual structures"
 /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript 3' end of cDNA cloned into XhoI site of pBluescript"

Query Match 0.7%; Score 21; DB 10; Length 343;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2915 atttggaatcctaagaact 2935
 Db 19 ATTTGAGAACTTACGAACT 39

RESULT 26
 AA821339 343 bp mRNA EST 17-FEB-1998
 LOCUS AA821339
 DEFINITION vs68a01.r1 Strataegene mouse skin (#937313) Mus musculus cDNA clone IMAGE:1151400 5', mRNA sequence.

ACCESSION AA821339
 VERSION AA821339.1 GI:2891207
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 343)
 AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Scheiellenberg,K., Stepcie,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
 The WashU-HMI Mouse EST Project
 Unpublished (1996)
 Contact: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.edu
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
 MGI:624608
 Seg primer: -28m13 rev1 ET from Amersham
 High quality sequence stop: 336.
 Location/Qualifiers
 1..343
 /organism="Mus musculus"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE:1151400"
 /clone_lib="Strataegene mouse skin (#937313)"
 /sex="females"
 /tissue_type="whole skin"
 /dev_stage="11 weeks old"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Organ: skin; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dr. Whole skin from 11 week old C57BL/6 female mice. Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGCGACGACG 3' ~3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3'"

BASE COUNT 84 a 103 c 95 g 61 t

ORIGIN
 source
 1..343
 /organism="Mus musculus"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE:1151400"
 /clone_lib="Strataegene mouse skin (#937313)"
 /sex="females"
 /tissue_type="whole skin"
 /dev_stage="11 weeks old"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Organ: skin; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dr. Whole skin from 11 week old C57BL/6 female mice. Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGCGACGACG 3' ~3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3'"

Query Match 0.7%; Score 21; DB 10; Length 343;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 787 atgaatcaagccccaagaact 807
 Db 319 ATGAATTCAGCCCGACGACT 339

RESULT 27
 A1645009 371 bp mRNA EST 29-APR-1999
 LOCUS A1645009
 DEFINITION vs68a01.y1 Strataegene mouse skin (#937313) Mus musculus cDNA clone IMAGE:1151400 5', mRNA sequence.

ACCESSION A1645009
 VERSION A1645009.1 GI:4723484
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 371)
 AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Stepcie,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
 The WashU-NCI Mouse EST Project 1999
 Unpublished (1999)
 Contact: Marra M/WashU-NCI Mouse EST Project 1999
 Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810

Email: mousest@watson.wustl.edu
This clone is available royalty-free through LINTL; contact the
IMAGE Consortium (info@image.liml.gov) for further information.
MG:624608

This read is a RESEQUENCE of a previously sequenced mouse clone
This read has been verified (found to hit its original self in the
correct orientation)

Seq primer: -40RP from Gibco

High quality sequence stop: 355.

FEATURES

SOURCE

1. .371

/organism="Mus musculus"

/strain="C57BL/6"

/db_xref="taxon:10090"

/clone="IMAGE:1151400"

/clone_lib="Stratagene mouse skin (#937313)"

/sex="females"

/tissue_type="whole skin"

/dev_stage="11 weeks old"

/lab_host="SOLR (kanamycin resistant)"

/note="Organ: skin; Vector: pBluescript SK-; Site_1: EcoRI

; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo

DT. Whole skin from 11 week old C57BL/6 female mice.

Average insert size: 1.0 kb; Uni-ZAP XR Vector: -5'

adaptor sequence: 5' GAATTCGGCAGAG 3' -3' adaptor

sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3'

BASE COUNT
ORIGIN

92 a 109 c 101 g 69 t

Query Match

Best Local Similarity 100.0%; Pred. No. 41;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY

787 atgaattcaagccagcagact 807

|||||

Db 321 ATGAATTCAGCCCGAGACT 341

RESULT 28

AZ223093/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Mouse BAC End Sequences from Library RPCI-23

Unpublished (1999)

Other GSS: RPCI-23-94H20.TJ

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC

library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)

or from Resea ch Genetics (Info@resgen.com). BAC end page:

http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 94 row: H column: 20
Seq primer: 17
Class: BAC ends.

FEATURES

SOURCE

1. .432

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPCI-23-94H20"

/clone_lib="RPCI-23"

/sex="female"

/lab_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site_1:

EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or

brain genomic DNA was isolated and partially digested

with a combination of EcoRI and EcoRI Methylase. Size

selected DNA was cloned into the pBAC3.6 vector at the

EcoRI sites. The ligation products were transformed into

DH10B electrocompetent cells (BRL Life Technologies).

BASE COUNT
ORIGIN

75 a 123 c 122 g 112 t

Query Match

Best Local Similarity 100.0%; Pred. No. 41;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY

787 atgaattcaagccagcagact 807

|||||

Db 115 ATGAATTCAGCCCGAGACT 95

RESULT 29

A0555443/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Sequence-tagged connectors: A sequence approach to mapping and

scanning the human genome

Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

99380589

Contact: Mahairas CG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPCI-11. For BAC

library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)

or from Resea ch Genetics (Info@resgen.com). BAC end Web Server:

http://www.hsc.washington.edu

Plate: 800 row: C column: 10

Seq primer: SP6

Class: BAC ends

High quality sequence stop: 437.

Location/Qualifiers

1. .437

/organism="Homo sapiens"

```

/db_xref="taxon:9606"
/clone="Plate=800 Col=10 Row=C"
/clone_lib="RPC1-11 Human Male BAC Library"
/sex="male"
/Note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRII. Size selected DNA was cloned into the
pBAC3.6 vector at EcoRI sites"
BASE COUNT      168 a      60 c      81 g      128 t
ORIGIN

Query Match      0.7%; Score 21; DB 13; Length 437;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1705 gtggcaccattatatacaag 1725
|||||
Db 387 gtgcacccattatatacaag 367

RESULT 30
CNS00SV2 462 bp DNA GSS 28-JUN-1999
CNS00SV2/c Arabidopsis thaliana genome survey sequence SP6 end of BAC T2N9 of
LOCUS TAMU library from strain Columbia of Arabidopsis thaliana, genomic
DEFINITION survey sequence.
ACCESSION AL088892.1 GI:5290032
VERSION AL088892.1
KEYWORDS tshale cress.
SOURCE Arabidopsis thaliana
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 462)
REFERENCE Salamon, D., Saurin, W., Artiguenave, F., Brotier, P., Wincker, P.,
AUTHORS Samoun, D., Saurin, W., Weissbach, J. and Queller, F.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 462)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
FEATURES
source
1.462
Location/Qualifiers
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone_lib="TAMU"
/clone="T2N9"
/Note="end : SP6"
BASE COUNT      129 a      87 c      70 g      176 t
ORIGIN

Query Match      0.7%; Score 21; DB 13; Length 462;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 626 caaatgtcttctctctcg 646
|||||
Db 239 CAAATGCTCTCTCTCCG 219

RESULT 31
A2360053 602 bp DNA GSS 02-OCT-2000
LOCUS 1M0103H11F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION clone UUGC1M0103H11 F, DNA sequence.
ACCESSION A2360053

```

```

VERSION A2360053.1 GI:10473753
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 602)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
M., Rose, M., Rose, R., Stokes, R., Tinney, A., von Niederhausern, A.
and Wright, D., Weiss, R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 1000 Std Error: 0.00
Plate: 0103 row: H column: 11
Seq primer: CGTTGTAAACGACGCGCAGT
Class: plasmid ends
High quality sequence stop: 602.
FEATURES
source
1.602
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="MGC1M0103H11"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/Note="Vector: PMD42ny; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PMD42 (g1147321149b1AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT      139 a      153 c      139 g      171 t
ORIGIN

Query Match      0.7%; Score 21; DB 13; Length 602;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1745 cctgagccagaatttgaagc 1765
|||||
Db 21 CCTGAGCCAGAAATTGAAC 1

RESULT 32
BG571228 1147 bp mRNA EST 10-APR-2001
LOCUS BG571228
DEFINITION 602592168F1 NIH_MGC_79 Homo sapiens cDNA clone IMAGE:4719424 5',
mRNA sequence.

```



```
/organism="Onchocerca volvulus"
/db_xref="taxon:6282"
/clone="SMOVL2CASI5B03"
/clone_lib="Onchocerca volvulus L2 larvae cDNA
(SAW98MLM-OvL2)"
/dev_stage="L2"
/lab_host="X11-Blue MRF"
/notes="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
Xho I; Filarial nematode parasite of humans. mRNA was
prepared from approximately 9,000 L2s isolated from
infected mosquitoes from Kumda, Cameroon and converted to
double-stranded cDNA using reverse transcriptase and
oligo(dT) followed by RNase H and DNA pol I. The library
has 7.3 x 10E4 independent recombinants and the average
insert size is approximately 1kb. The library was
constructed by Michelle Lizotte-Waniewski. The library is
available from Dr. S.A. Williams, email: genomesmith.edu."

BASE COUNT      22 a      11 c      19 g      21 t
ORIGIN

Query Match      0.6%; Score 20; DB 10; Length 73;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2364 gaagaagatcgtataaact 2383
|||||
Db 27 GAAGAAATCTCTATAAACT 46

RESULT 35
LOCUS      AO069062      205 bp      DNA      GSS      04-AUG-1998
DEFINITION HS-2255-B2_H03_MR CIT Approved Human Genomic Sperm Library D Homo
ACCESSION  AO069062
VERSION     AO069062.1 GI:3384261
KEYWORDS    GSS.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 205)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 2255 Row: P Column: 6
Class: BAC ends
High quality sequence stop: 205.
Location/Qualifiers
1. 205
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2255 Col=6 Row=P"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/notes="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
E-Coli DH10B"

BASE COUNT      54 a      41 c      52 g      57 t      1 others
ORIGIN
```

```
Query Match      0.6%; Score 20; DB 13; Length 205;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1147 tacagaacaacacacacaa 1166
|||||
Db 134 TACAGAAACAAACACAA 115

RESULT 36
LOCUS      B1315256      215 bp      mRNA      EST      20-JUL-2001
DEFINITION dah96e06.y1 NICHD XGC Emb4 Xenopus laevis cDNA clone IMAGE:4957906
5', mRNA sequence.
ACCESSION  B1315256
VERSION     B1315256.1 GI:14989583
KEYWORDS    EST.
SOURCE      African clawed frog.
Xenopus laevis
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 215)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Martha Rebert, Steven L. Klein, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: Xenopus clones from this library are available
through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
Seq primer: -40RP from Gibco
High quality sequence stop: 197.
Location/Qualifiers
1. 215
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="IMAGE:4957906"
/clone_lib="NICHD XGC Emb4"
/dev_stage="embryo, stage 31-32"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: whole embryo; Vector: PCMV-SPORT6; Site_1:
NotI; Site_2: SalI; Cloned unidirectionally. Primer: oligo
dT. Average insert size 2.1 kb. Constructed by life
Technologies. Note: This is a Xenopus Gene Collection (XGC
) library."

BASE COUNT      72 a      32 c      37 g      74 t
ORIGIN

Query Match      0.6%; Score 20; DB 11; Length 215;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2825 ttctggaagaacctctga 2844
|||||
Db 196 TTTTGAAGAACACCTCTGA 215

RESULT 37
LOCUS      AA493806      216 bp      mRNA      EST      19-AUG-1997
DEFINITION nb02f04.s1 NCI_CGAP_Thy1 Homo sapiens cDNA clone IMAGE:943135, mRNA
sequence.
ACCESSION  AA493806
VERSION     AA493806.1 GI:2223647
KEYWORDS    EST.
SOURCE      human.
```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 216)
NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Kitzman, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone Distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LINM at:
www-bio.lnlnl.gov/bdrp/image/image.html
Insert Length: 294 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham.

FEATURES
source
1. 216
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:943135"
/clone_lib="NCI-CCAP_Thy1"
/tissue_type="thyroid"
/lab_host="DH10B"
/note="Vector: PAMP10; mRNA made from invasive thyroid
tumor, CDNA made by oligo-dT priming. Non-directionally
cloned. Size selected on agarose gel, average insert size
600 bp. Reference: Kitzman et al. (1996) Cancer Research
56:5380-5383."

BASE COUNT 57 a 65 c 37 g 57 t
ORIGIN

Query Match 0.68; Score 20; DB 10; Length 216;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2225 gcaatcacatctgaacaa 2244
|||||
Db 92 GCACATCAGCTGTACCA 111

RESULT 38
AI366637 225 bp mRNA EST 08-JAN-1999
LOCUS SMOV3MCAM11C04SK Onchocerca volvulus molting L3 larva CDNA
DEFINITION (SI96MLW-Ovml3) Onchocerca volvulus CDNA clone SMOV3MCAM11C04 5',
mRNA sequence.
ACCESSION AI366637
VERSION AI366637.1 GI:4136382
KEYWORDS EST.
SOURCE Onchocerca volvulus.
ORGANISM Onchocerca volvulus.
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
Onchocercidae; Onchocerca.
REFERENCE 1 (bases 1 to 225)
Williams, S.A., Lizotte-Waniewski, M., Laney, S. and Lustigman, S.
Genes expressed in molting L3 larvae of Onchocerca volvulus
Unpublished (1997)
AUTHORS Contact: Steven A. Williams
JOURNAL Molecular Parasitology
COMMENT Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genome@smith.edu
Seq primer: pBluescript SK.
Location/Qualifiers

FEATURES
source

source
1. 225
/organism="Onchocerca volvulus"
/strain="Kumba, Cameroons"
/db_xref="taxon:6282"
/clone="SMOV3MCAM11C04"
/clone_lib="Onchocerca volvulus molting L3 larva CDNA
(SI96MLW-Ovml3)"
/dev_stage="molting L3"
/lab_host="XLI-Blue MRF"
/note="Vector: Lambda Uni-Zap XR; Site 1: Eco RI; Site 2:
Xho I; Filarial nematode parasite of humans, third-stage
larvae, L3, were isolated from infected black flies in
Cameroon (forest strain). The L3 were cultured in 20% FCS
in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in
culture. L3 of O. volvulus molt to fourth-stage larvae by
day 5 in culture. mRNA was isolated from approximately
6000 molting larvae (ml3) 2000 larvae from day 1, 2 or 3
in culture, and converted to double-stranded cDNA using
reverse transcriptase and oligo(dT) followed by RNase H
and DNA pol I. The library was constructed in the lambda
Uni-Zap XR vector and has 1 x 10E6 independent
recombinants and the average insert size is ~1200 bp. The
library was constructed by Sara Lustigman and Michelle
Lizotte-Waniewski in the Laboratory of Dr. S. A. Williams.
The library is available from Dr. Sara Lustigman (email:
slustig@nybc.org)."

BASE COUNT 65 a 39 c 57 g 64 t
ORIGIN

Query Match 0.68; Score 20; DB 10; Length 225;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2364 gaagaagatgctataaact 2383
|||||
Db 30 GAAGAAGATGCTATAA ACT 49

RESULT 39
AI603832 248 bp mRNA EST 21-APR-1999
LOCUS SMOV3MCAM27A08SK Onchocerca volvulus molting L3 larva CDNA
DEFINITION (SI96MLW-Ovml3) Onchocerca volvulus CDNA clone SMOV3MCAM27A08 5',
mRNA sequence.
ACCESSION AI603832
VERSION AI603832.1 GI:4612981
KEYWORDS EST.
SOURCE Onchocerca volvulus.
ORGANISM Onchocerca volvulus.
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
Onchocercidae; Onchocerca.
REFERENCE 1 (bases 1 to 248)
Williams, S.A., Lizotte-Waniewski, M., Laney, S. and Lustigman, S.
Genes expressed in molting L3 larvae of Onchocerca volvulus
Unpublished (1997)
AUTHORS Contact: Steven A. Williams
JOURNAL Molecular Parasitology
COMMENT Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genome@smith.edu
Seq primer: pBluescript SK.
Location/Qualifiers

FEATURES
source
1. 248
/organism="Onchocerca volvulus"
/strain="Kumba, Cameroons"
/db_xref="taxon:6282"
/clone="SMOV3MCAM27A08"
/clone_lib="Onchocerca volvulus molting L3 larva CDNA
(SI96MLW-Ovml3)"


```

/dev-stage="molting L3"
/lab_host="XLI-Blue MRF"
/Note="Vector: Lambda Uni-ZAP XR. Site_1: Eco RI; Site_2:
Xho I; Filarial nematode parasite of humans. Third-stage
larvae, L3, were isolated from infected black flies in
Cameroun (forest strain). The L3 were cultured in 20% FCS
in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in
culture. L3 of O. volvulus molt to fourth-stage larvae by
day 5 in culture. mRNA was isolated from approximately
6000 molting larvae (ml3), 2000 larvae from day 1, 2 or 3
in culture, and converted to double-stranded cDNA using
reverse transcriptase and oligo(dT) followed by RNase H
and DNA pol I. The library was constructed in the lambda
Uni-Zap XR vector and has 1 x 10E6 independent
recombinants and the average insert size is ~1200 bp. The
library was constructed by Sara Lustigman and Michelle
Lizotte-Manlewski in the Laboratory of Dr. S. A. Williams.
The library is available from Dr. Sara Lustigman (email:
slustig@nybc.org)."
BASE COUNT      71 a      47 c      58 g      69 t      3 others
ORIGIN
Query Match      0.6%; Score 20; DB 10; Length 248;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2364 gaagaagatgctataaact 2383
|||||
Db 14 GAAGAAGATGCTATMAAACT 33

RESULT 40
LOCUS      AZ891870      299 bp      DNA      GSS      05-MAR-2001
DEFINITION RPCI-24-180D14.TJ RPCI-24 Mus musculus genomic clone RPCI-24-180D14
ACCESSION  AZ891870
VERSION     AZ891870
KEYWORDS    GSS.
SOURCE      house mouse.
ORGANISM    Mus musculus
REFERENCE   1 (bases 1 to 299)
AUTHORS     Zhao,S., Niernan,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Georegeorgis,E.,
Russell,D., de Jong,P. and Fraser,C.M.
TITLE       Mouse BAC End Sequences from Library RPCI-24
JOURNAL     Unpublished (1999)
COMMENT     Other GSSs: RPCI-24-180D14.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szha@etigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@etigr.org). Clones may be purchased from BACPAC
Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
page: http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html
Plate: 180 row: D column: 14
Seq primer: SP6
Class: BAC ends.
FEATURES
Source      1..299
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="RPCI-24-180D14"
/clone_lib="RPCI-24"

```

```

/sex="Male"
/cell_type="Spleen/Brain"
/Note="Vector: pPARBAC1. Site_1: BamHI; Site_2: BamHI;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pPARBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."
BASE COUNT      103 a      57 c      85 g      54 t
ORIGIN
Query Match      0.6%; Score 20; DB 13; Length 299;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 989 tgaagctgtgtctccaaa 1008
|||||
Db 202 TGAAGCTGTGTCTCCAAA 221

RESULT 41
LOCUS      AW357412      332 bp      mRNA      EST      25-APR-2001
DEFINITION 40862 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION  AW357412
VERSION     AW357412.1 GI:6861490
KEYWORDS    EST.
SOURCE      cow.
ORGANISM    Bos taurus
REFERENCE   1 (bases 1 to 332)
AUTHORS     Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,D., Cho,J., Fahrnenkrug,S.C., Bennett,
G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
Perte,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
Keeler,J.W.
TITLE       Sequence evaluation of four pooled-tissue normalized bovine cDNA
JOURNAL     libraries and construction of a gene index for cattle
MEDLINE    Genome Res. 11 (4), 626-630 (2001)
COMMENT     Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and trimmed with phred
v0.980904.e. Vector identified by cross_match with the -mnscore 20
and -mismatch 12 options.
PCR Primers
FORWARD: AGGAACACCTATGACCAT
BACKWARD: GTTTCACGACGACGACG
Plate: 20 row: J column: 17
Seq primer: ATTGAGTGACACTATAG.
FEATURES
Source      1..332
Location/Qualifiers
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 3BOV"
/tissue_type="pooled"
/lab_host="DH10b"
/Note="Vector: PCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from marrow, alveolar
macrophage, ovary, fetal semitendinosus muscle, and fetal
longissimus muscle."
BASE COUNT      59 a      105 c      88 g      80 t
ORIGIN
Query Match      0.6%; Score 20; DB 10; Length 332;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;

```

Matches 20: Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 733 tggccatgctgctgaagctg 752
 |||||||
 Db 77 TGGCCATGCTGCTGAAGCTG 96

RESULT 42

A0952434/c

LOCUS 336 bp DNA GSS 27-JAN-2000
 DEFINITION Shared DNA-48N4.TR Sheared DNA Trypanosoma brucei genomic clone

ACCESSION A0952434
 VERSION A0952434.1 GI:6775699

KEYWORDS GSS.
 SOURCE Trypanosoma brucei
 ORGANISM Trypanosoma brucei

TITLE Trypanosoma. Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
 JOURNAL Unpublished (1999)

REFERENCE 1 (bases 1 to 336)
 El-Sayed, N., Zhao, H., Gill, S., Suh, E., Malek, J., Fujii, C.,
 Gerrard, C., Leech, V., de Jong, P., Ullu, E., Melville, S., Donelson, J.,
 Fraser, C. and Adams, M.

COMMENT Determination of clone end sequences from Trypanosoma brucei GUTat
 10.1 sheared DNA library
 Other_GSSs: Sheared DNA-48N4.TF
 Contact: Najib M. El-Sayed
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208

Email: nelsayed@tigr.org
 Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
 DNA library constructed at TIGR. Clones will be available for
 distribution through ATCC. Sheared DNA end sequences search page:
 http://www.tigr.org/tldb/mdb/tldb/.
 Seq primer: M13-Reverse
 Class: shotgun.

FEATURES
 Location/Qualifiers
 1..336
 /organism="Trypanosoma brucei"
 /strain="TREU927/4 GUTat 10.1"
 /db_xref="taxon:5691"
 /clone="Sheared DNA-48N4"
 /clone_lib="Sheared DNA"

/note="Vector: pUC18; Site_1: SmaI; Constructed at The
 Institute for Genomic Research (TIGR), Rockville, MD.
 Genomic DNA isolated from a cloned population of
 Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
 sheared to give a tight size distribution (approx 2 kb).
 The v + i method used for the library construction is
 described in detail in Smith, H.O. and Venter, J.C.
 (Making small insert libraries for whole genome shotgun
 sequencing projects. In Genome Sequencing: A Practical
 Approach, eds. M. Yaudin and B. Bartell, Oxford University
 Press, 1999)."

BASE COUNT 72 a 62 c 68 g 134 t
 ORIGIN

Query Match 0.6%; Score 20; DB 13; Length 336;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 20: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1153 aaacaacacacataaa 1172
 |||||||
 Db 171 AAAACAACACAAACATAA 152

RESULT 43
 AA670222/c

LOCUS AA670222 342 bp mRNA EST 20-NOV-1997
 DEFINITION ad19hl.s1 Soares_NBHF Homo sapiens cDNA clone IMAGE:878757 3',
 mRNA sequence.
 ACCESSION AA670222
 VERSION AA670222.1 GI:2631721
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 342)

AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
 Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Matra, M., Matlin
 White, Y., Wylie, T., Waterston, R. and Wilson, R.

TITLE WashU-NCI human EST Project
 JOURNAL Unpublished (1997)
 CONTACT: Wilson RK

COMMENT Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810

Email: est@wustl.edu
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -40m13 fwd. ET from Amersham
 High quality sequence stop: 334.

FEATURES
 Location/Qualifiers
 1..342

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:878757"
 /clone_lib="Soares_NBHF"
 /dev_stage="15 wk post natal"
 /lab_host="MDH10B"
 /note="Organ: whole brain; Vector: p7773D-Pac (Pharmacia)
 with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
 1st strand cDNA was primed with a Not I - oligo(dT) primer
 [5' AACGTGAGAGATTCGGCGCGCAATATTTTATTTTATTTT 3']
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified p7773 vector. Library
 went through one round of normalization. Library
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 72 a 122 c 81 g 67 t
 ORIGIN

Query Match 0.6%; Score 20; DB 10; Length 342;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 20: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 648 ctccacagggcccgaggtcg 667
 |||||||
 Db 137 CTCACAGGCGCCACAGGTGG 118

RESULT 44

A2991565

LOCUS 380 bp DNA GSS 27-APR-2001
 DEFINITION 2M0275B24R Mouse 10kb plasmid U06C2M library Mus musculus genomic
 clone U06C2M0275B24 R, DNA sequence.

ACCESSION A2991565
 VERSION A2991565.1 GI:13862792

KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 380)
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
 M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.

TITLE and Wright, D., Weiss, R.
JOURNAL Mouse whole genome scaffolding with paired end reads from 10kb
COMMENT plasmid inserts
 Unpublished (2000)
 Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0275 row: B column: 24
 Seq primer: CACACGGAACACGATGACC
 Class: plasmid ends
 High quality sequence stop: 380.
 Location/Qualifiers
 1. 380

FEATURES

source
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="U0GC2M0275B24"
 /clone_lib="Mouse 10kb plasmid U0GC2M library"
 /sex="Female"
 /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g147321149b1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 140 a 71 c 97 g 72 t
ORIGIN

Query Match 0.6%; Score 20; DB 13; Length 380;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1153 aaacaacaacaacataaa 1172
 ||||||||||||||||||||
Db 35 AAAACAACACCAACATVAA 54

RESULT 45
AA294641 393 bp mRNA EST 12-NOV-1997
LOCUS SMOV3MCA885SK Onchocerca volvulus molting L3 larva cDNA
DEFINITION (SL96MLM-Ovml3) Onchocerca volvulus cDNA clone Sml3C0885 5', mRNA sequence.
ACCESSION AA294641
VERSION AA294641.1 GI:2099635
KEYWORDS EST.
SOURCE Onchocerca volvulus.
ORGANISM Onchocerca volvulus.
 Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea; Onchocercidae; Onchocerca.
REFERENCE 1 (bases 1 to 393)
AUTHORS Williams, S.A., Lizotte-Waniewski, M., Laney, S. and Lustigman, S.

TITLE Genes expressed in molting L3 larvae of Onchocerca volvulus
JOURNAL Unpublished (1997)
COMMENT Contact: Steven A. Williams
 Molecular Parasitology
 Smith College Department of Biological Sciences
 Department of Biological Sciences, Clark Science Center, Smith
 College, Northampton, MA, 01063, USA
 Tel: 4135853826
 Fax: 4135853786
 Email: genome@smith.edu
 Seq primer: pBluescript SK.
 Location/Qualifiers
 1. 393

FEATURES

source
 /organism="Onchocerca volvulus"
 /strain="Kumba, Cameroons"
 /db_xref="taxon:6282"
 /clone="Sml3C0885"
 /clone_lib="Onchocerca volvulus molting L3 larva cDNA (SL96MLM-Ovml3)"
 /dev_stage="molting L3"
 /lab_host="XLI-Blue MRF"
 /note="Vector: Lambda Uni-ZAP XR; Site-1: Eco RI; Site-2: Xho I; Filarial nematode parasite of humans. Third-stage larvae, L3, were isolated from infected black flies in Cameroon (forest strain). The L3 were cultured in 20% FCS in IMDM+NCIC 135 and collected after day 1, 2, or 3 in culture. L3 of O. volvulus molt to fourth-stage larvae by day 5 in culture. mRNA was isolated from approximately 6000 molting larvae (ml3). 2000 larvae from day 1, 2 or 3 in culture, and converted to double-stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H and DNA pol I. The library was constructed in the lambda Uni-Zap XR vector and has 1 x 10⁶ independent recombinants and the average insert size is ~1200 bp. The library was constructed by Sara Lustigman and Michelle Lizotte-Waniewski in the Laboratory of Dr. S. A. Williams. The library is available from Dr. Sara Lustigman (email: slustigman@ncpc.org)."
BASE COUNT 104 a 79 c 96 g 111 t 3 others
ORIGIN

Query Match 0.6%; Score 20; DB 10; Length 393;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2364 gaagaagatgctataaaact 2383
 ||||||||||||||||||||
Db 42 GAAGAAGATGCTATAAAACT 61

RESULT 46
AM838955 393 bp mRNA EST 18-MAY-2000
LOCUS CM2-LT0061-180200-094-h05 LT0061 Homo sapiens cDNA, mRNA sequence.
DEFINITION AM838955
ACCESSION AM838955
VERSION AM838955.1 GI:7932929
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 393)
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663

COMMENT

Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?lib=st2-cm2-LR0061-180
200-094-105<3=2000-02-18<4=1)
Seq primer: puc 18 forward
High quality sequence stop: 393.

FEATURES

source

```
1..393
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="LR0061"
/dev_stage="Adult"
/note="Organ: leiomyos; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT      90 a      99 c      115 g      89 t
ORIGIN
```

Query Match 0.6%; Score 20; DB 10; Length 393;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1815 tactattgactctctga 1834

Db 68 TACTTATTGACTCTTCTTGA 49

RESULT 47

AA294540

LOCUS 415 bp mRNA EST 12-NOV-1997

DEFINITION SMOV3MCA61SK Onchocerca volvulus molting L3 larva cDNA

(SI96MLW-ovml3) Onchocerca volvulus cDNA SMOV3C0761 5', mRNA

Sequence.

ACCESSION AA294540

VERSION AA294540.1

KEYWORDS GI:2099534

SOURCE EST.

ORGANISM Onchocerca volvulus.

Onchocerca volvulus

Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;

Onchocercidae; Onchocerca.

1 (bases 1 to 415)

Williams, S.A., Lizotte-Waniewski, M., Laney, S. and Lustigman, S.

Genes expressed in molting L3 larvae of Onchocerca volvulus

Unpublished (1997)

Contact: Steven A. Williams

Molecular Parasitology

Smith College Department of Biological Sciences

Department of Biological Sciences, Clark Science Center, Smith

College, Northampton, MA, 01063, USA

Tel: 4135853826

Fax: 4135853786

Email: genome@smith.edu

Seq primer: PBluescript SK.

Location/Qualifiers

1..415

/organism="Onchocerca volvulus"

/strain="Kumba, Camerouns"

/db_xref="taxon:6283"

/clone="SMV3C0761"

/clone_lib="Onchocerca volvulus molting L3 larva cDNA"

(SI96MLW-ovml3)"

/dev_stage="molting L3"

/lab_host="XLI-Blue MRF"

/note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:

Xho I; Filarial nematode parasite of humans. Third-stage

larvae, L3, were isolated from infected black flies in

Cameron (Forest strain). The L3 were cultured in 20% FCS

in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in

culture. L3 of O. volvulus molt to fourth-stage larvae by

day 5 in culture. mRNA was isolated from approximately

6000 molting larvae (ml3). 2000 larvae from day 1, 2 or 3

in culture, and converted to double-stranded cDNA using

reverse transcriptase and oligo(dT), followed by RNase H

and DNA pol I. The library was constructed in the Lambda

Uni-Zap XR vector and has 1 x 10⁶ independent

recombinants and the average insert size is ~1200 bp. The

library was constructed by Sara Lustigman and Michelle

Lizotte-Waniewski in the laboratory of Dr. S. A. Williams.

The library is available from Dr. Sara Lustigman (email:

slustigman@bc.org)."

BASE COUNT 110 a 89 c 103 g 102 t 11 others

ORIGIN

Query Match 0.6%; Score 20; DB 10; Length 415;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2364 gaagaagatgctataaact 2383

Db 17 GACAGACATGCTATTAACCT 36

RESULT 48

T81898/c

LOCUS 415 bp mRNA EST 15-MAR-1995

DEFINITION yd29a11.s1 Soares fetal liver spleen 1NF1S Homo sapiens cDNA clone

IMAGE:109628 3', mRNA sequence.

ACCESSION T81898

VERSION T81898.1

KEYWORDS GI:704905

SOURCE EST.

ORGANISM human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 415)

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman

, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,

Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston

, R., Williamson, A., Wollmann, P. and Wilson, R.

The WashU-Merck EST Project

Unpublished (1995)

Other ESTs: yd29a11.r1

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Insert size: 714

High quality sequence stops: 314

This clone is available royalty-free through LBNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 714 Std Error: 0.00

Seq primer: T3

High quality sequence stop: 314.

Location/Qualifiers

1..415

/organism="Homo sapiens"

/db_xref="GDB:465245"

/db_xref="taxon:9606"

/clone="IMAGE:109628"

/clone_lib="Soares fetal liver spleen 1NF1S"

LOCUS BF601894 491 bp mRNA EST 25-APR-2001
DEFINITION 266990 MARC 3BOV Bos taurus CDNA 5', mRNA sequence.
ACCESSION BF601894
VERSION BF601894.1 GI:11699117
KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrnenkrug,S.C., Bennett,
'G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
Pettea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
Keeler,J.W.
Sequence evaluation of four pooled-tissue normalized bovine CDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
21180013
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt. trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACACGCTATGACCAT
BACKWARD: GTTTCACAGTCACGACG
Plate: 42 row: N column: 11
Seq primer: ATTAGGTGACACTATAG.
FEATURES
source
1. .491
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_id="MARC 3BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; site_1: XbaI; site_2: XhoI;
library made from pooled tissue from marrow, alveolar
macrophage, ovary, fetal semitendinosus muscle, and fetal
longissimus muscle."
BASE COUNT 120 a 104 c 105 g 162 t
ORIGIN

Query Match 0.6%; Score 20; DB 11; Length 491;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2787 agactcacagatcacagat 2806
|||||
DB 284 AGACTCACAGATCACAGAT 265

RESULT 52
LOCUS A1643174 498 bp mRNA EST 29-APR-1999
DEFINITION V446909.Y1 Soares_mammary_gland_NbMWG Mus musculus CDNA clone
IMAGE:864448 5', mRNA sequence.
ACCESSION A1643174
VERSION A1643174.1 GI:4721649
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: ccgaps-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:508536
This read is a RESEQUENCE of a previously sequenced mouse clone
correct orientation)
Seq primer: -40RP from Glibco
High quality sequence stop: 484.
FEATURES
source
1. .498
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:864448"
/clone_id="Soares_mammary_gland_NbMWG"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia
) with a modified polylinker; Site_1: Not I; Site_2: Eco
RI; 1st strand CDNA was primed with a Not I - oligo(dT)
primer [5',
TGTATCCAACTGTGAAGTGGGAGCGCGCGAGATGTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded CDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M. Fatima
Bonaldo."
BASE COUNT 138 a 130 c 108 g 121 t 1 others
ORIGIN

Query Match 0.6%; Score 20; DB 10; Length 498;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1327 cagctcaagggttcagcca 1346
|||||
DB 498 CAGCTCAAGGTTCAGCCA 479

RESULT 53
LOCUS BF602580 505 bp mRNA EST 25-APR-2001
DEFINITION 267845 MARC 3BOV Bos taurus CDNA 5', mRNA sequence.
ACCESSION BF602580
VERSION BF602580.1 GI:11699804
KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrnenkrug,S.C., Bennett,
'G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
Pettea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
Keeler,J.W.
Sequence evaluation of four pooled-tissue normalized bovine CDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
21180013
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366

Fax: 402 762 4390
 Email: smith@email.marc.usda.gov
 Single pass sequencing. Bases called and alt. trimmed with phred
 v0.980904.e. Vector identified by cross_match with the -minscore 18
 and -mismatch 12 options.

PCR Primers
 FORWARD: AGGAACAGCTATGACCAT
 BACKWARD: GTTTCACAGTCACGACG
 Plate: 44 row: K column: 14
 Seq primer: ATTTAGTGCACCTATAG.

FEATURES

Source

Location/Qualifiers

1. 505

/organism="Bos taurus"

/db_xref="taxon:9913"

/clone_lib="MARC 3BOV"

/tissue_type="pooled"

/lab_host="DH10B"

/note="Vector: PCMV SPORT6; Site_1: XbaI; Site_2: XhoI;

Library made from pooled tissue from marrow, alveolar

macrophage, ovary, fetal semitendinosus muscle, and fetal

longissimus muscle."

BASE COUNT 96 a 155 c 132 g 122 t

ORIGIN

Query Match 0.6%; Score 20; DB 11; Length 505;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 733 tggccatcgtctgaaagctg 752

Db 174 TGGCAGTCGTCGACACTG 193

RESULT 54
 BG382487/c 519 bp mRNA EST 12-MAR-2001
 LOCUS 298398 MARC 1P1G Sus scrofa CDNA 5', mRNA sequence.
 DEFINITION BG382487
 ACCESSION BG382487.1 GI:13306959
 VERSION EST.
 KEYWORDS
 SOURCE Sus scrofa
 ORGANISM pig.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
 1 (bases 1 to 519)
 Fahnenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
 Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
 and Keele,J.W.
 Design and use of two pooled tissue normalized CDNA libraries for
 EST discovery in swine
 Unpublished (2000)
 Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@email.marc.usda.gov
 Single pass sequencing. Bases called and alt. trimmed with phred
 v0.980904.e. Vector identified by cross_match with the -minscore 18
 and -mismatch 12 options.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

Source

Location/Qualifiers

1. 519

/organism="Sus scrofa"

/db_xref="taxon:9823"

/clone_lib="MARC 1P1G"

/tissue_type="pooled"

/lab_host="DH10B"

/note="Vector: PCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
 Library made from pooled tissue from day 11, 13, 15, 20,
 and 30 embryos "

BASE COUNT 150 a 103 c 104 g 162 t

ORIGIN

Source

Location/Qualifiers

1. 521

/organism="Onchocerca volvulus"

/db_xref="taxon:6282"

/clone_lib="SMOV3MCM26G12"

/dev_stage="molting L3"

/lab_host="XLI-Blue MRP"

/note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:

Xho I; Filarial nematode parasite of humans. Third-stage

larvae, L3, were isolated from infected black flies in

Cameroon (forest strain). The L3 were cultured in 20% FCS

in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in

culture. L3 of O. volvulus molt to fourth-stage larvae by

day 5 in culture. mRNA was isolated from approximately

6000 molting larvae (ml3), 2000 larvae from day 1, 2 or 3

in culture, and converted to double-stranded cDNA using

reverse transcriptase and oligo(dT) followed by RNase H

and DNA pol I. The library was constructed in the lambda

Uni-Zap XR vector and has 1 x 10⁶ independent

recombinants and the average insert size is ~1200 bp. The

library was constructed by Sara Lustigman and Michelle

Lizotte-Maniowski in the laboratory of Dr. S. A. Williams.

The library is available from Dr. Sara Lustigman (email:

slustigman@dc.org)."

BASE COUNT 139 a 105 c 125 g 147 t 5 others

ORIGIN

RESULT 55
 A1670486 521 bp mRNA EST 17-MAY-1999
 LOCUS SMOV3MCM26G12SK Onchocerca volvulus molting L3 larva CDNA
 DEFINITION (SL96MLM-OvML3) Onchocerca volvulus cDNA clone SMOV3MCM26G12 5',
 mRNA sequence.
 ACCESSION A1670486
 VERSION A1670486.1 GI:4836992
 KEYWORDS EST.
 SOURCE Onchocerca volvulus.
 ORGANISM Onchocerca volvulus.
 Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
 Onchocercidae; Onchocerca.
 1 (bases 1 to 521)
 Williams,S.A., Lizotte-Maniowski,M., Laney,S. and Lustigman,S.
 Genes expressed in molting L3 larvae of Onchocerca volvulus
 Unpublished (1997)
 Contact: Steven A. Williams
 Molecular Parasitology
 Smith College Department of Biological Sciences
 Department of Biological Sciences, Clark Science Center, Smith
 College, Northampton, MA, 01063, USA
 Tel: 4135853826
 Fax: 4135853786
 Email: genomemsmith.edu
 Seq primer: pbuescript SK.
 Location/Qualifiers

FEATURES

Source

Location/Qualifiers

1. 521

/organism="Onchocerca volvulus"

/db_xref="taxon:6282"

/clone_lib="SMOV3MCM26G12"

/dev_stage="molting L3"

/lab_host="XLI-Blue MRP"

/note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:

Xho I; Filarial nematode parasite of humans. Third-stage

larvae, L3, were isolated from infected black flies in

Cameroon (forest strain). The L3 were cultured in 20% FCS

in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in

culture. L3 of O. volvulus molt to fourth-stage larvae by

day 5 in culture. mRNA was isolated from approximately

6000 molting larvae (ml3), 2000 larvae from day 1, 2 or 3

in culture, and converted to double-stranded cDNA using

reverse transcriptase and oligo(dT) followed by RNase H

and DNA pol I. The library was constructed in the lambda

Uni-Zap XR vector and has 1 x 10⁶ independent

recombinants and the average insert size is ~1200 bp. The

library was constructed by Sara Lustigman and Michelle

Lizotte-Maniowski in the laboratory of Dr. S. A. Williams.

The library is available from Dr. Sara Lustigman (email:

slustigman@dc.org)."

Query Match 0.6%; Score 20; DB 10; Length 521;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2364 gaagaagatgcataaact 2383
 ||||||||||||||||
 DB 42 GAAGAAGATGCTATAAACT 61

RESULT 56
 FR0031506/c
 LOCUS 521 bp DNA GSS 27-JUN-1998
 DEFINITION Fugu rubripes GSS sequence, clone 116B21aA4, genomic survey
 sequence.
 ACCESSION AL027875.1 GI:3269989
 VERSION AL027875.1
 KEYWORDS GSS; genome survey sequence.
 SOURCE Takifugu rubripes.
 ORGANISM Takifugu rubripes.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Acanthopterygii; Neopterygii; Teleostei; Euteleostei;
 Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 Tetraodontidae; Takifugu.
 1 (bases 1 to 521)
 Elgar, G., Clark, M., Smith, S., Meek, S., Warner, S., Umranta, Y.,
 Williams, G. and Brenner, S.
 TITLE Direct Submission
 JOURNAL Submitted (09-JUN-1998) MRC Human Genome Mapping Project Resource
 Centre, Hinxton, Cambridgeshire, CB10 1SB, UK. Email:
 biohelp@hgm.mrc.ac.uk
 Vector: pBluescript II KS
 V-type: phagemid
 PRIMER: KS
 DESCR: One pass dye-terminator sequencing of cosmid cloned genomic
 sequence.

FEATURES
 source Location/Qualifiers
 1..521
 /organism="Takifugu rubripes"
 /db_xref="taxon:31033"
 /clone_lib="cosmid 116B21"
 /clone="116B21aA4"

BASE COUNT 125 a 120 c 120 g 136 t 20 others

ORIGIN

Query Match 0.6%; Score 20; DB 13; Length 521;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2620 aagatggaagaagctctt 2639
 ||||||||||||||||
 DB 312 AAGATGAAGTGAAGCTCTT 293

RESULT 57
 BI394563
 LOCUS 526 bp mRNA EST 06-AUG-2001
 DEFINITION pcp1n.pk001.110 Normalized Chicken Pituitary/Hypothalamus/Pineal
 Library Gallus gallus cDNA clone pcp1n.pk001.110 5' similar to
 g117710042 ref1np.057930.11 IO motif containing GTPase activating
 protein 1; Cdc42-Rac1 effector protein [Mus musculus]
 sp|Q9URF1|IOG1.MOUSE RAS GTPASE-ACTIVATING-LIKE PROTEIN IOGAPI
 gblAA660344.1|AF240630.1 (AF240630) IO motif containing GTPas, mRNA
 sequence.
 ACCESSION BI394563
 VERSION BI394563.1 GI:15087845
 KEYWORDS EST.
 SOURCE chicken.
 ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.

REFERENCE 1 (bases 1 to 526)
 AUTHORS Porter, T.E. and Cogburn, L.A.
 TITLE ESTs from Normalized Chicken Pituitary/Hypothalamus/Pineal cDNA
 Library USDA/IRAPs Animal Genome Project
 JOURNAL Unpublished (2001)
 COMMENT Contact: Larry A. Cogburn
 University of Delaware
 Townsend Hall, Newark, DE 19717, USA
 Tel: 302-831-1335
 Fax: 302-831-2822
 Email: cogburnudel.edu, www.chickest.udel.edu.

FEATURES
 source Location/Qualifiers
 1..526
 /organism="Gallus gallus"
 /strain="Commercial broiler chicken"
 /db_xref="taxon:9031"
 /clone="pcp1n.pk001.110"
 /clone_lib="Normalized Chicken
 Pituitary/Hypothalamus/Pineal Library"
 /sex="Male and Female"
 /tissue_type="Pituitary Gland/Hypothalamus/Pineal Gland"
 /dev_stage="Embryonic (d12,d14,d19); post-hatch (w1,w3,w5
 w7,w9)."
 /lab_host="E. Coli EMDH10B"
 /note="Vector: pCMVSPORT6; Library made from equivalent
 pools of total RNA isolated from each tissue at different
 ages. Single pass sequencing from 5'-end"
 BASE COUNT 124 a 162 c 151 g 87 t 2 others

ORIGIN

Query Match 0.6%; Score 20; DB 11; Length 526;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 740 gctgctgaagctgcgcaga 759
 ||||||||||||||||
 DB 72 GCTGCTGAAGCTGCGCAGA 91

RESULT 58
 BG351255
 LOCUS 535 bp mRNA EST 01-MAR-2001
 DEFINITION 08BD06 Mature tuber lambda ZAP Solanum tuberosum cDNA, mRNA
 sequence.
 ACCESSION BG351255
 VERSION BG351255.1 GI:13179997
 KEYWORDS EST.
 SOURCE potato.
 ORGANISM Solanum tuberosum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
 1 (bases 1 to 535)
 Nielsen, K.L., Crookshanks, M., Emmersen, J. and Wellinder, K.G.
 EST-sequencing of mature potato tuber (Var. Kurat)
 Unpublished (2000)
 Contact: Karen G. Wellinder
 Institut for bioteknologi
 Aalborg Universitet
 Sohngaardssholmsvej 49, 9000 Aalborg, Denmark
 Tel: +45 96358467
 Fax: +45 98141808
 Email: kgw@bio.auc.dk
 Sequenced from the 5' end.
 High quality sequence stop: 535
 POLYA-No.

FEATURES
 source Location/Qualifiers
 1..535
 /organism="Solanum tuberosum"
 /cultivar="Field grown Kurat"
 /db_xref="taxon:4113"
 /clone_lib="Mature tuber lambda ZAP"

BASE COUNT 177 a 110 c 109 g 139 t
 ORIGIN /tissue_type="Tuber"
 /note="Vector: Lambda ZAP"

Query Match 0.6%; Score 20; DB 11; Length 535;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2395 tgaataactgaagaagatg 2414
 |||||||
 Db 160 TGAATAAAGCTGAGAGATG 179

LOCUS 59
 221490 A221490 541 bp DNA GSS 14-JUN-2000

DEFINITION Gm_UMB001_125_D12.R UMN Soybean BAC Library (PECSBAC4 EcORI)
 Glycine max genomic clone Gm_UMB001_125_D12, DNA sequence.

ACCESSION A221490
 VERSION A221490.1 GI:8517274
 KEYWORDS GSS.

SOURCE soybean.
 ORGANISM Glycine max

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseolae;
 Glycine.

1 (bases 1 to 541)
 AUTHORs Marek, L.F., Paz, M., Darnielle, L., Hanson, N. and Shoemaker, R.C.
 TITLE BAC End sequences from a soybean genomic library (ISU)
 JOURNAL Unpublished (2000)
 COMMENT Contact: Shoemaker Randy C
 Agronomy Department
 Iowa State University
 Ames, IA 50011-1010, USA
 Tel: 515 294 1205
 Fax: 515 294 2299
 Email: rcschoe@iastate.edu

FEATURES
 source This BAC identified by SSR Satt427. For more information, see
 SoyBase at:
 http://genome.cornell.edu/cgi-bin/WebAce/webacetdb-soybase.

Please see as an authority for the mapping/naming: Cregan P.B., T.
 Jarvik, A.L. Bush, R.C. Shoemaker, K.G. Lark, A.L. Kahler, N. Kaya,
 T.T. Vantaoi, D.G. Lohnes, J. Chung, and J.E. Specht. 1999a. An
 integrated genetic linkage map of the soybean genome. Crop Sci.
 39: (in press)
 Seq primer: BACR or M13R
 Class: BAC ends.

FEATURES
 source Location/Qualifiers

1..541
 /organism="Glycine max"
 /cultivar="Faribault"
 /db_xref="taxon:3847"
 /clone="Gm_UMB001_125_D12"
 /clone_lib="UMN Soybean BAC Library (PECSBAC4 EcORI)"
 /tissue_type="cotyledon leaves"
 /dev_stage="cotyledon stage"
 /note="Vector: PECSBAC4; The UMN BAC library (Danesh et al
 /Theor. Appl. Genet. 96:196, 1998) was constructed using
 the Eco RI site of PECSBAC4. The library consists of 72
 ,960 clones with an average insert size of 120 kb, equal
 to 7 haploid genome equivalents. Screening of the library
 is done by PCR amplification of DNA pools."
 BASE COUNT 177 a 88 c 90 g 165 t 21 others
 ORIGIN

Query Match 0.6%; Score 20; DB 13; Length 541;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1024 gcttgagaatcatcatgaag 1043
 |||||||
 Db 424 GCTTGAGAACTGCATCATGAG 443

RESULT 60
 A2298434 554 bp DNA GSS 27-JUL-2000
 LOCUS RPCI-23-109A14.TV RPCI-23 Mus musculus genomic clone RPCI-23-109A14
 DEFINITION , DNA sequence.

ACCESSION A2298434
 VERSION A2298434.1 GI:9540219
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 554)
 AUTHORs Zhao, S., Niernan, W., Feldblyum, T., Malek, J., Shatsman, S., Akinret,
 B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.
 and Fraser, C.M.
 TITLE Mouse BAC End Sequences from Library RPCI-23
 JOURNAL Unpublished (1999)
 COMMENT Other GSSs: RPCI-23-109A14.TJ
 Contact: Shaying Zhao
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC
 library availability, please contact Pieter de Jong
 (pieter@edlong.med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
 or from Resea ch Genetics (info@resgen.com). BAC end page:
 http://www.tigr.org/cdb/bac_ends/mouse/bac_end_intro.html
 Plate: 109 row: A column: 14
 Seq primer: T7
 Class: BAC ends.

FEATURES
 source Location/Qualifiers

1..554
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-23-109A14"
 /clone_lib="RPCI-23"
 /sex="Female"
 /lab_host="DH10B"
 /note="Organ: Kidney/Brain; Vector: pBACe3.6; site.1:
 EcORI; site.2: EcORI; Female: C57BL/6J mouse kidney and/or
 brain genomic DNA was isolated and partially digested
 with a combination of EcORI and EcORI Methylase. Size
 selected DNA was cloned into the pBACe3.6 vector at the
 EcORI sites. The ligation products were transformed into
 DH10B electrocompetent cells (BRL Life Technologies)."
 BASE COUNT 179 a 80 c 76 g 219 t
 ORIGIN

Query Match 0.6%; Score 20; DB 13; Length 554;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1868 cattaactggaactttatg 1887
 |||||||
 Db 97 CATTAAGCTGAGACTTTATG 116

RESULT 61
 A0270247 557 bp DNA GSS 03-NOV-1998
 LOCUS HS-2045_A1_G04_MR CIT Approved Human Genomic Sperm Library D Homo
 DEFINITION

```

ACCESSION      sapiens genomic clone Plate=2045 Col=7 Row=M, DNA sequence.
VERSION        AQ270247.1 GI:3822842
KEYWORDS       GSS.
SOURCE         human.
ORGANISM       Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS        Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
              1 (bases 1 to 557)
              Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
              Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
              Hood,L.
              Sequence-tagged connectors: A sequence approach to mapping and
              scanning the human genome
              Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
JOURNAL        99380589
MEDLINE        Contact: Mahairas GG, Wallace JC, Hood L
COMMENT        High Throughput Sequencing Center
              University of Washington
              401 Queen Anne Avenue North, Seattle, WA 98109, USA
              Tel: (206) 616-3618
              Fax: (206) 616-3887
              Email: jwallace@u.washington.edu
              Sequence Tagged Connector
              Plate: 2045 Row: M Column: 7
              Class: BAC ends
              High quality sequence stop: 557.
              Location/Qualifiers
                1..557
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="Plate=2045 Col=7 Row=M"
                /clone_1lb="CIT Approved Human Genomic Sperm Library D"
                /sex="male"
                /note="Organ: sperm; Vector: pBelBAC11; BAC Clones in
                E-Coli DH10B"
BASE COUNT     137 a 140 c 108 g 169 t 3 others
ORIGIN
Query Match    0.6%; Score 20; DB 13; Length 557;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1940 catggaagagggcccaagaa 1959
|||||
Db 192 CATGGAAGAGGCCCGAGAA 173

RESULT 62
BF258009/c     571 bp mRNA EST 23-FEB-2001
LOCUS          HVSMEF0014H19f Hordeum vulgare seedling root EST library HVCDNM0007
DEFINITION     (etiolated and unstressed) Hordeum vulgare cDNA clone
ACCESSION      BF258009
VERSION        BF258009.2 GI:13119185
KEYWORDS       EST.
SOURCE         barley.
ORGANISM       Hordeum vulgare
REFERENCE      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS        Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
              1 (bases 1 to 571)
              Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Fritsch,D., Yu
              X., Anderson,H., Dale,J., Henry,D., Kennolly,S., Palmer,M., Rambo
              W.T., Saski,C., Schwartzbeck,J., Simmons,J., Choi,D.W., Main,D. and
              Wood,T.
              Development of a genetically and physically anchored EST resource
              for barley genomics
              Unpublished (2000)
              On Nov 16, 2000 this sequence version replaced gi:11187122.
              Contact: Wing RA

```

```

Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: AATTAACTTCACATGAAGG
High quality sequence stop: 558.
Location/Qualifiers
  1..571
  /organism="Hordeum vulgare"
  /cultivar="Morex"
  /db_xref="taxon:4513"
  /clone="HVSMEF0014H19f"
  /clone_1lb="Hordeum vulgare seedling root EST library
  HVCDNM0007 (etiolated and unstressed)"
  /tissue_type="Seedling root"
  /lab_host="TJC121"
  /note="Vector: lambdaZAP; site_1: EcoRI; site_2: XhoI; For
  more details on library preparation and sequence analysis
  see http://www.genome.clemson.edu/projects/barley/ To
  order a clone see http://www.genome.clemson.edu/orders"
BASE COUNT     101 a 166 c 218 g 86 t
ORIGIN
Query Match    0.6%; Score 20; DB 11; Length 571;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 500 tggcctcctgcagctcttc 519
|||||
Db 191 TGGCCTCCTGCAGCCTCTTC 172

RESULT 63
BG385206/c     577 bp mRNA EST 12-MAR-2001
LOCUS          BG385206
DEFINITION     306845 MARC 1Pig Sus scrofa cDNA 5', mRNA sequence.
ACCESSION      BG385206
VERSION        BG385206.1 GI:13309678
KEYWORDS       EST.
SOURCE         pig.
ORGANISM       Sus scrofa
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS        Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
              1 (bases 1 to 577)
              Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
              Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
              and Keeler,J.W.
              Design and use of two pooled tissue normalized cDNA libraries for
              EST discovery in swine
              Unpublished (2000)
              Contact: Smith TPL
              USDA, ARS, US Meat Animal Research Center
              PO Box 166, Clay Center, NE 68933-0166, USA
              Tel: 402 762 4366
              Fax: 402 762 4390
              Email: smithhemall.marc.usda.gov
              Single pass sequencing. Bases called and alt_trimmed with phred
              v0.980904.e. Vector identified by cross_match with the -minscore 18
              and -mismatch 12 options.
              PCR primers
              FORWARD: AGGAACAGCATATGACCAT
              BACKWARD: GTTTTCCAGTCACGACG
              Plate: 94 Row: J Column: 10
              Seg primer: ATTAGTGACATATAG.
              Location/Qualifiers
                1..577
                /organism="Sus scrofa"
                /db_xref="taxon:9823"
                /clone_1lb="MARC 1Pig"
                /tissue_type="pooled"

```

/lab_host="DH10B"
/note="Vector: PCMV SPORF6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from day 11, 13, 15, 20,
and 30 embryos."
BASE COUNT 102 a 200 c 175 g 100 t
ORIGIN

Query Match 0.6%; Score 20; DB 11; Length 577;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 420 aactgaagaagcaccctcac 439
|||||
Db 572 AACTGAAAGCACCTTCAC 553

RESULT 64
LOCUS A0570315 577 bp DNA GSS 01-JUN-1999
DEFINITION HS.5352.B1.E02.T7A RPCI-11 Human Male BAC Library Homo sapiens
genomic clone Plate-928 Col=3 Row=J, DNA sequence.
ACCESSION A0570315
VERSION A0570315.1 GI:4963535
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 577)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589

TITLE High Throughput Sequencing Center
JOURNAL Contact: Mahairas GG, Wallace JC, Hood L
MEDLINE University of Washington
COMMENT 401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.bufileo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.bufileo.edu/ordering_bac.htm)
or from Research Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 928 row: J column: 3
Seq primer: 17
Class: BAC ends
High quality sequence stop: 577.
Location/Qualifiers

FEATURES
source 1..577
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate-928 Col=3 Row=J"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: PBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
PBACE3.6 vector at EcoRI sites"
BASE COUNT 244 a 114 c 82 g 130 t 7 others
ORIGIN

Query Match 0.6%; Score 20; DB 13; Length 577;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2105 caggtcgaataaagat 2124
|||||
Db 348 CAGGCTGCAATAAAGAGAT 367

RESULT 65
LOCUS A2987277/c 580 bp DNA GSS 27-APR-2001
DEFINITION 2M0269J15R Mouse 10kb plasmid UUGC2M library Mus musculus genomic
clone UUGC2M0269J15 R, DNA sequence.
ACCESSION A2987277
VERSION A2987277.1 GI:13858504
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 580)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
CONTACT: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0269 row: J column: 15
Seq primer: CACACGACAAACAGCATGACC
Class: Plasmid ends
High quality sequence stop: 580.
Location/Qualifiers

TITLE Mouse whole genome scaffolding with paired end reads from 10kb
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0269 row: J column: 15
Seq primer: CACACGACAAACAGCATGACC
Class: Plasmid ends
High quality sequence stop: 580.
Location/Qualifiers

FEATURES
source 1..580
Location/Qualifiers

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0269J15"
/clone_lib="Mouse 10kb plasmid UUGC2M library"
/sex="Female"
/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PMD42ny; Purified genomic DNA from M.
musculus C57BL/6J (female) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (9114732114[9b]AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT 169 a 91 c 93 g 227 t
ORIGIN

Query Match 0.6%; Score 20; DB 13; Length 580;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1338 ttcaagccaagataaatt 1357
 Db 479 TTCAAGCCCAAGTATAAATT 460

RESULT 66

LOCUS BG098273

DEFINITION BG098273 605 bp mRNA EST 29-JAN-2001
 EST462792 sprouting eyes/shoots Solanum tuberosum cDNA clone
 C57C2015 5' sequence, mRNA sequence.

ACCESSION BG098273

VERSION BG098273.1 GI:12586308

KEYWORDS EST.

SOURCE Solanum tuberosum

ORGANISM Solanum tuberosum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
 1 (bases 1 to 605)
 van der Hoeven,R.S., Bezzerides,J., Cho,J., Uterback,T., Hansen
 ,C.L., Bougri,O., Buell,C.R., Ronning,C.M., Tanksley,S.D. and Baker

REFERENCE

AUTHORS

TITLE Generation of ESTs from potato sprouting eyes/shoots
 JOURNAL Unpublished (2001)
 COMMENT Contact: Cathy Ronning
 The Institute for Genomic Research
 For clone info: please contact Research Genetics, Libraries
 Division tel 1-800-711-6195, email cdna@resgen.com.

FEATURES

Source

1..605

/organism="Solanum tuberosum"

/cultivar="Kennebec"

/db_xref="taxon:4113"

/clone="C57C2015"

/clone_lib="sprouting eyes/shoots"

/cruise_type="sprouting tubers"

/dev_stage="12-14 weeks post harvest"

/lab_host="SOLR"

/note="Vector: pBluescript SK(-); Site1: EcoRI; Site2:
 XhoI: Various sizes of sprouting eyes (2 mm to 15 mm) were
 taken from tubers. The tubers were incubated at 26C in
 the dark for 2-3 weeks prior to sprouting. The eyes were
 frozen in liquid nitrogen immediately upon removal from
 tubers."

BASE COUNT 192 a 117 c 119 g 177 t

ORIGIN

Query Match 0.6%; Score 20; DB 11; Length 605;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2395 tgaataacctgaagaagt 2414

156 TGAATAACCTGAAGAAGTG 175

RESULT 67

LOCUS A2802117

DEFINITION A2802117 620 bp DNA GSS 16-FEB-2001
 2M0061H04F Mouse 10kb plasmid U06C1M library Mus musculus genomic
 clone U06C2M0061H04 F, DNA sequence.

ACCESSION A2802117

VERSION A2802117.1 GI:12954536

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 620)
 Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,

TITLE

JOURNAL

COMMENT

Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
 ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.
 and Wright,D., Weiss,R.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 Unpublished (2000)

CONTACT: Robert B. Weiss

University of Utah Genome Center

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0061 row: H column: 04

Seq primer: CGTGTAAACGACGCCACAT

Class: plasmid ends

High quality sequence stop: 620.

FEATURES

Source

1..620

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="U06C2M0061H04"

/clone_lib="Mouse 10kb plasmid U06C1M library"

/sex="Male"

/lab_host="E. coli strain XL10-Gold, TI-resistant, F-"

/note="Vector: PMD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adapted DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of PMD42 (91473211419b/AP129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adapted mouse DNA was annealed to
 adapted vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

BASE COUNT 167 a 141 c 175 g 137 t

ORIGIN

Query Match 0.6%; Score 20; DB 13; Length 620;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2145 ctcaagttgtctcagcac 2164

180 CTCAGTTGTCTCAGCAC 199

RESULT 68

LOCUS A1511575

DEFINITION A1511575 627 bp mRNA EST 16-MAR-1999
 SMOVL3CAN30H03SK Onchocerca volvulus infective larva cDNA
 (SAM94WL-Ov13) Onchocerca volvulus cDNA SMOVL3CAN30H03 5',
 mRNA sequence.

ACCESSION A1511575

VERSION A1511575.1 GI:4417429

KEYWORDS EST.

SOURCE Onchocerca volvulus.

ORGANISM Onchocerca

Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
 Onchocercidae; Onchocerca.

REFERENCE 1 (bases 1 to 627)
AUTHORS Williams, S.A., Lu, W., Lizotte-Maniewski, M. and Laney, S.J.
TITLE Genes expressed in infective third stage larvae of *Onchocerca volvulus*
JOURNAL Unpublished (1995)
COMMENT Contact: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genome@smith.edu
Seq primer: pbluescript SK.
Location/Qualifiers
1. 627
/organism="Onchocerca volvulus"
/strain="Sierra Leone"
/db_xref="taxon:6282"
/clone="SMOVL3CAN30H03"
/clone_lib="Onchocerca volvulus infective larva cDNA (SAM94WL-OVL3)"
/lab_host="XLI-Blue MRF"
/note="Vector: lambda UniZap XR; Site 1: EcoR I; Site 2: Xho I; Cutaneous filarial nematode parasite of humans. *Onchocerca volvulus* isolated from mosquitoes 10 days after infection and converted to double stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H and DNase I. The library had 1.8 x 10⁵ independent recombinants and average insert size was 900 base pairs. The library was constructed by Wenhong Lu. The library is available from Dr. S.A. Williams, email genome@smith.edu."

BASE COUNT 173 a 123 c 148 g 183 t
ORIGIN

Query Match 0.6%; Score 20; DB 10; Length 627;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2364 gaagaagatgcataaaact 2383
|||||
b 28 GAGGAGATGCTATAAACT 47

RESULT 69
A2510047 633 bp DNA GSS 05-OCT-2000
LOCUS 1M0354H14F Mouse 10kb plasmid UUGCIM library Mus musculus genomic
DEFINITION clone UUGCIM0354H14 F, DNA sequence.
ACCESSION A2510047
VERSION A2510047.1 GI:10691363
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Mus.
1 (bases 1 to 633)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A. and Wright, D., Weiss, R.
Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177

Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0354 row: H column: 14
Seq primer: CGTTGTAACGACGGCCACT
Class: Plasmid ends
High quality sequence stop: 633.
Location/Qualifiers
1. 633
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGCIM0354H14"
/clone_lib="Mouse 10kb plasmid UUGCIM library"
/sex="Male"
/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PMD42ny; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (g114732114|g9|AF12072.1), a copy-number inducible derivative of plasmid RI. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 147 a 154 c 134 g 198 t
ORIGIN

Query Match 0.6%; Score 20; DB 13; Length 633;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 707 tggcacatcaggaagcaga 726
|||||
Db 257 TGGCACATCAGGAGCAGA 276

RESULT 70
AM874886 644 bp mRNA EST 22-MAY-2000
LOCUS SMOVNFCA10G04SK Onchocerca volvulus microfilaria cDNA
DEFINITION (SAM98MLM-OVMF) Onchocerca volvulus clone SMOVNFCA10G04 5', mRNA sequence.
ACCESSION AM874886
VERSION AM874886.1 GI:8012581
KEYWORDS EST.
SOURCE Onchocerca volvulus.
ORGANISM Onchocerca volvulus.
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea; Onchocercidae; Onchocerca.
1 (bases 1 to 644)
Williams, S.A.
Genes expressed in microfilaria of *Onchocerca volvulus*
Unpublished (1999)
Contact: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genome@smith.edu
Seq primer: pbluescript SK.

FEATURES
SOURCE

location/Qualifiers
1.644
/organism="Onchocerca volvulus"
/db_xref="taxon:6282"
/clone="SMOV3MCA07B09"
/clone_1lb="Onchocerca volvulus microfilaria cDNA (SAM98MLM-OvMF)"
/dev_stage="microfilaria"
/lab_host="XLI-Blue MRF"
/note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2: Xho I; Filarial nematode parasite of humans. mRNA was prepared from approximately 200,000 microfilariae isolated from the skin of infected individuals from Kumba, Cameroon and converted to double-stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H and DNA pol I. The library has 7.8 x 10E4 independent recombinants and the average insert size is approximately 1kb. The library was constructed by Michelle Lizotte-Waniewski. The library is available from Dr.S.A.Williams, email:genomesmith.edu."

BASE COUNT
ORIGIN
182 a 125 c 152 g 185 t

Query Match
Best Local Similarity 100.0%; Score 20; DB 10; Length 644;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2364 gaagaagatgctataaact 2383
|||||
Db 27 GAAGAAGATGCTATTAACACT 46

RESULT 71
BE346038 652 bp mRNA EST 17-JUL-2000
LOCUS SMOV12CASA10G04SK Onchocerca volvulus L2 larvae cDNA (SAM98MLM-OvL2)
DEFINITION Onchocerca volvulus cDNA clone SMOV12CASA10G04 5', mRNA sequence.
ACCESSION BE346038
VERSION BE346038.1 GI:9255574
KEYWORDS EST.
SOURCE Onchocerca volvulus.
ORGANISM Onchocerca volvulus.
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
Onchocercidae; Onchocerca.
REFERENCE
1 (bases 1 to 652)
Williams, S.A.
Genes expressed in L2 larvae of Onchocerca volvulus
Unpublished (1999)
COMMENT
Contact: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genomesmith.edu
Seq primer: pBluescript SK.
Location/Qualifiers
1.652
/organism="Onchocerca volvulus"
/db_xref="taxon:6282"
/clone="SMOV12CASA10G04"
/clone_1lb="Onchocerca volvulus L2 larvae cDNA (SAM98MLM-OvL2)"
/dev_stage="L2"
/lab_host="XLI-Blue MRF"
/note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2: Xho I; Filarial nematode parasite of humans. mRNA was prepared from approximately 9,000 L2s isolated from infected mosquitoes from Kumba, Cameroon and converted to double-stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H and DNA pol I. The library

has 7.3 x 10E4 independent recombinants and the average insert size is approximately 1kb. The library was constructed by Michelle Lizotte-Waniewski. The library is available from Dr.S.A.Williams, email: genomesmith.edu."

BASE COUNT
ORIGIN
187 a 124 c 157 g 184 t

Query Match
Best Local Similarity 100.0%; Score 20; DB 10; Length 652;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2364 gaagaagatgctataaact 2383
|||||
Db 27 GAAGAAGATGCTATTAACACT 46

RESULT 72
A1317901 653 bp mRNA EST 17-DEC-1998
LOCUS SMOV3MCA07B09SK Onchocerca volvulus molting L3 larva cDNA
DEFINITION (SL96MLM-OvML3) Onchocerca volvulus cDNA clone SMOV3MCA07B09 5', mRNA sequence.
ACCESSION A1317901
VERSION A1317901.1 GI:4033168
KEYWORDS EST.
SOURCE Onchocerca volvulus.
ORGANISM Onchocerca volvulus.
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
Onchocercidae; Onchocerca.
REFERENCE
1 (bases 1 to 653)
Williams, S.A., Lizotte-Waniewski, M., Laney, S. and Lustigman, S.
Genes expressed in molting L3 larvae of Onchocerca volvulus
Unpublished (1997)
COMMENT
Contact: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genomesmith.edu
Seq primer: pBluescript SK.
Location/Qualifiers
1.653
/organism="Onchocerca volvulus"
/strain="Kumba, Cameroons"
/db_xref="taxon:6282"
/clone="SMOV3MCA07B09"
/clone_1lb="Onchocerca volvulus molting L3 larva cDNA (SL96MLM-OvML3)"
/dev_stage="molting L3"
/lab_host="XLI-Blue MRF"
/note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2: Xho I; Filarial nematode parasite of humans. Third-stage larvae, L3, were isolated from infected black flies in Cameroon (forest strain). The L3 were cultured in 20% FCS in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in culture. L3 of O. volvulus molt to fourth-stage larvae by day 5 in culture. mRNA was isolated from approximately 6000 molting larvae (mJ3), 2000 larvae from day 1, 2 or 3 in culture, and converted to double-stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H and DNA pol I. The library was constructed in the lambda Uni-Zap XR vector and has 1 x 10E5 independent recombinants and the average insert size is ~1200 bp. The library was constructed by Sara Lustigman and Michelle Lizotte-Waniewski in the Laboratory of Dr. S. A. Williams. The library is available from Dr. Sara Lustigman (email: slustigman@bc.org)."

BASE COUNT
ORIGIN
184 a 127 c 160 g 182 t

Query Match 0.6%; Score 20; DB 10; Length 653;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2364 gaagaagatgctataaact 2383
 Db 31 GAAGAAGATGCTATAAAACT 50

RESULT 73

BE636495

LOCUS 656 bp mRNA EST 25-AUG-2000
 DEFINITION SMOVL2CASI7D04SK Onchocerca volvulus L2 larvae cDNA (SAM98MLM-OVL2)

ONCHOCERCA VOLVULUS CDNA clone SMOVL2CASI7D04 5', mRNA sequence.

BE636495

KEYWORDS GI:9919606

EST. Onchocerca volvulus.

SOURCE Onchocerca volvulus.

ORGANISM Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;

REFERENCE 1 (bases 1 to 656)

AUTHORS Williams, S.A.

TITLE Genes expressed in L2 larvae of Onchocerca volvulus

JOURNAL Unpublished (1999)

COMMENT Contact: Steven A. Williams

Molecular Parasitology Smith College Department of Biological Sciences

Department of Biological Sciences, Clark Science Center, Smith

College, Northampton, MA, 01063, USA

Tel: 4135853826

Fax: 4135853786

Email: genome@smith.edu

Seq primer: pBluescript SK.

Location/Qualifiers

1. 656

/organism="Onchocerca volvulus"

/db_xref="taxon:6282"

/clone="SMOVL2CASI7D04"

/clone_1lb="Onchocerca volvulus L2 larvae cDNA

(SAM98MLM-OVL2)"

/dev_stage="L2"

/lab_host="XLI-Blue MRF"

/note="Vector: Lambda Uni-Zap XR; Site_1: Eco RI; Site_2:

Xho I; Filarial nematode parasite of humans. mRNA was

prepared from approximately 9,000 L2s isolated from

infected mosquitoes from Kumba, Cameroon and converted to

double-stranded cDNA using reverse transcriptase and

oligo(dT) followed by RNase H and DNA pol I. The library

has 7.3 x 10B4 independent recombinants and the average

insert size is approximately 1kb. The library was

constructed by Michelle Lizotte-Waniewski. The library is

available from Dr. S.A. Williams, email: genome@smith.edu."

BASE COUNT 183 a 129 c 157 g 186 t 1 others

ORIGIN

Query Match 0.6%; Score 20; DB 10; Length 656;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2364 gaagaagatgctataaact 2383

Db 28 GAAGAAGATGCTATAAAACT 47

RESULT 74

LOCUS 660 bp mRNA EST 23-FEB-2001

DEFINITION BG310513

SMOV3MCAM56C02SK Onchocerca volvulus molting L3 larva cDNA

(SL96MLM-OVML3) Onchocerca volvulus cDNA clone SMOV3MCAM56C02 5',

mRNA sequence.

ACCESSION BG310513
 VERSION BG310513.1 GI:13112371
 KEYWORDS EST.
 SOURCE Onchocerca volvulus.
 ORGANISM Onchocerca volvulus.

REFERENCE 1 (bases 1 to 660)
 Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
 Onchocercidae; Onchocerca.

AUTHORS Williams, S.A., Lizotte-Waniewski, M., Laney, S. and Lustigman, S.
 Genes expressed in molting L3 larvae of Onchocerca volvulus

Unpublished (1997)

COMMENT Contact: Steven A. Williams

Molecular Parasitology Smith College Department of Biological Sciences

Department of Biological Sciences, Clark Science Center, Smith

College, Northampton, MA, 01063, USA

Tel: 4135853826

Fax: 4135853786

Email: genome@smith.edu

Seq primer: pBluescript SK.

Location/Qualifiers

1. 660

/organism="Onchocerca volvulus"

/strain="Kumba, Cameroons"

/db_xref="taxon:6282"

/clone="SMOV3MCAM56C02"

/clone_1lb="Onchocerca volvulus molting L3 larva cDNA

(SL96MLM-OVML3)"

/dev_stage="molting L3"

/lab_host="XLI-Blue MRF"

/note="Vector: Lambda Uni-Zap XR; Site_1: Eco RI; Site_2:

Xho I; Filarial nematode parasite of humans. Third-stage

larvae, L3, were isolated from infected black flies in

Cameroon (forest strain). The L3 were cultured in 20% FCS

in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in

culture. L3 of O. volvulus molt to fourth-stage larvae by

day 5 in culture. mRNA was isolated from approximately

6000 molting larvae (mL3). 2000 larvae from day 1, 2 or 3

in culture, and converted to double-stranded cDNA using

reverse transcriptase and oligo(dT) followed by RNase H

and DNA pol I. The library was constructed in the lambda

Uni-Zap XR vector and has 1 x 10B6 independent

recombinants and the average insert size is ~1200 bp. The

library was constructed by Sara Lustigman and Michelle

Lizotte-Waniewski in the Laboratory of Dr. S. A. Williams.

The library is available from Dr. Sara Lustigman (email:

slustig@nhybc.org)."

BASE COUNT 187 a 127 c 161 g 185 t

ORIGIN

Query Match 0.6%; Score 20; DB 11; Length 660;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2364 gaagaagatgctataaact 2383

Db 28 GAAGAAGATGCTATAAAACT 47

RESULT 75

LOCUS 726 bp DNA GSS 12-MAY-2000

DEFINITION CENS02935

Tetradon nigroviridis genome survey sequence PUC-01 end of clone

247A05 of library G from Tetradon nigroviridis, genomic survey

sequence.

ACCESSION AL186746

VERSION AL186746.1 GI:7824850

KEYWORDS GSS: genome survey sequence.

SOURCE Tetradon nigroviridis.

ORGANISM Tetradon nigroviridis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.

REFERENCE 1 (bases 1 to 726)
AUTHORS Roest-Crolius,H., Jallou,O., Dasilva,C., Fizames,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 726)
AUTHORS Roest-Crolius,H., Jallou,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 726)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
<http://www.genoscope.cns.fr/Tetraodon>.

FEATURES
SOURCE location/Qualifiers
1..726
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="247A05"
/clone_lib="G"
/note="Genoscope sequence ID : C0AG247A03SP1-end :
PUC-Or1"

BASE COUNT 231 a 162 c 139 g 185 t 9 others
ORIGIN

Query Match 0.6%; Score 20; DB 13; Length 726;
Best Local Similarity 100.0%; Pred.No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 215 aaaggttcagagtcctgta 234
|||||
DB 533 AAAGGTTTCAGAGTCCTGTA 514

Search completed: March 25, 2002, 15:27:50
Job time: 13639 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2002, 11:39:43 ; Search time 97.58 Seconds

(without alignments)
7129.949 Million cell updates/sec

Title: US-09-697-089-3

Perfect score: 3072

Sequence: 1 atgaattcctaagacaa.....ctttaactagtaactgct 3072

Scoring table: IDENTITY-NUC
Gapop 10.0 , Gapext 1.0

Archived: 351203 seqs, 11323899 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued Patents-NA:*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCITUS.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	59	1.9	5502	3	US-08-836-134-1
2	39.6	1.3	2862	4	US-08-569-749-13
3	39.6	1.3	2862	5	PCR-US96-12860-13
4	39.6	1.3	3151	3	US-09-212-971-13
5	39.6	1.3	3151	3	US-08-800-929A-13
6	39.6	1.3	3151	4	US-09-617-053A-13
7	36.8	1.2	1435	4	PCR-US95-05922A-1
8	36.8	1.2	2589	5	US-08-569-749-1
9	36.8	1.2	2589	5	PCR-US96-12860-1
10	36.8	1.2	3532	2	US-09-205-204-1
11	36.8	1.2	3732	3	US-09-212-971-7
12	36.8	1.2	3732	3	US-08-800-929A-7
13	36.8	1.2	3732	4	US-09-617-053A-7
14	35.2	1.1	2580	2	US-08-511-485-7
15	34.8	1.1	1840	2	US-08-484-993B-15
16	34.8	1.1	1840	2	US-08-484-156B-15
17	34.8	1.1	1840	2	US-08-484-596A-15
18	34.8	1.1	1840	2	US-08-480-150A-15
19	34.8	1.1	1840	3	US-08-458-731-15
20	34.8	1.1	1840	3	US-08-149-223A-15
21	34.8	1.1	53526	3	US-08-658-136-2
22	34.8	1.1	53577	3	US-08-658-136-1
23	34.2	1.1	510	2	US-08-934-959-7
24	34.2	1.1	7218	1	US-08-232-465-14
25	34	1.1	200	2	US-08-875-972-27
26	34	1.1	1753	6	5225348-2
27	34	1.1	4695	6	5225348-3

28	33.8	1.1	292	1	US-08-644-664B-41	Sequence 41, Appl
29	33.8	1.1	292	2	US-08-761-277A-41	Sequence 41, Appl
30	33.8	1.1	3979	4	US-09-180-439-1	Sequence 1, Appl
31	33.8	1.1	3979	4	US-09-180-439-2	Sequence 2, Appl
32	33.8	1.1	4123	4	US-09-180-439-7	Sequence 7, Appl
33	33.4	1.1	893	4	US-09-142-320-3	Sequence 3, Appl
34	33.4	1.1	893	4	US-09-142-320-8	Sequence 8, Appl
35	33.4	1.1	1359	2	US-08-851-088-1	Sequence 1, Appl
36	33.4	1.1	3573	4	US-09-353-585-4	Sequence 4, Appl
37	33.4	1.1	5535	1	US-08-089-755A-1	Sequence 1, Appl
38	33.4	1.1	5535	1	US-08-089-755A-4	Sequence 4, Appl
39	33.4	1.1	5535	1	US-08-421-754-1	Sequence 1, Appl
40	33.4	1.1	5535	1	US-08-421-754-4	Sequence 4, Appl
41	33.4	1.1	5535	2	US-08-421-791-1	Sequence 1, Appl
42	33.4	1.1	5535	2	US-08-421-791-4	Sequence 4, Appl
43	33.4	1.1	6471	4	US-09-353-585-1	Sequence 1, Appl
44	33.2	1.1	3624	1	US-07-951-715A-6	Sequence 6, Appl
45	33.2	1.1	3624	2	US-08-459-448A-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1

US-08-836-134-1

Sequence 1, Application US/08836134A

Patent No. 6020127

GENERAL INFORMATION:

APPLICANT: Mackenzie, Alex E.

APPLICANT: Korneluk, Robert G.

APPLICANT: Mahadevan, Mani S.

APPLICANT: McLean, Michael

APPLICANT: Roy, Natalie

APPLICANT: Ikeda, John

TITLE OF INVENTION: Neuronal Apoptosis Inhibitor Protein, Gene Sequence and

Patent No. 6020127

TITLE OF INVENTION: Mutations Causative of Spinal Muscular Atrophy

FILE REFERENCE: 3477-112, 033477/139914

CURRENT APPLICATION NUMBER: US/08/836,134A

CURRENT FILING DATE: 1997-06-20

NUMBER OF SEQ ID NOS: 23

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 1

LENGTH: 5502

TYPE: DNA

ORGANISM: Homo sapiens

US-08-836-134-1

Query Match

Best Local Similarity 54.3%; Score 59; DB 3; Length 5502;

Matches 144; Conservative 0; Mismatches 115; Indels 6; Gaps 1;

Qy	500	ttbaaggaagatctggcaagagcaagtcacactgtctgacgagcattgcatgctctggg	559
Db	1798	tggaggggttaagctgtaagtggaagacggtctctctgaagaatagctttctctggg	1857
Qy	560	gtctcggaagtgcaagctctgaccaagttcaattcgtctctctctcgtctcag--	617
Db	1858	catctgcatgctgtccctgttaacaggttccagctgtgtttcttaccctctccctagtt	1917
Qy	618	----caggcccaagtgtaacttttgaacacctctgtatacaacctccgataaccg	673
Db	1918	ccaccagaccagcaggggctgccaagatcatctgttaccacagctccctagagaagaag	1977
Qy	674	gcaacatcggaagcagacattcatgcatgctgttgtaagctgacgagaggtctctt	733
Db	1978	gactgttcttgtaagatgcatgaggaacatatccagcagttaaagaatcaggtcttat	2037
Qy	734	tcctcttgatgctacaaatgaatt	758
Db	2038	tccttttagatgactacaagaat	2062

```
RESULT 2
US-08-569-749-13
; Sequence 13, Application US/08569749
; Patent No. 6187557
; GENERAL INFORMATION:
; APPLICANT: Rothe, Mike
; APPLICANT: Goeddel, David V
; TITLE OF INVENTION: INHIBITORS OF APOPTOSIS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/569,749
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Brezner, David J.
; REGISTRATION NUMBER: 24,774
; REFERENCE/DOCKET NUMBER: A-62464/DJB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)781-1989
; TELEFAX: (415)398-3249
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2862 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-569-749-13

Query Match 1.3%; Score 39.6; DB 4; Length 2862;
Best Local Similarity 46.9%; Pred. No. 0.15;
Matches 123; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

QY 50 tgactgttataaagcaatacacagatgacctatttgatggaatgctcgaatcggaag 109
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1869 TGACACATGTCCTTCCTGATATCTCTTGAGGCCAGTGTAATTAACAAACAGG 1928
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 110 aagtaaacatcattgtcgagaagtgagcagagatgctgctagaaggatcattaca 169
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1929 AACATGATATTATTAGACAGAAAAACACAGATACCTTACAGCAGAGAGACTTTTGACA 1988
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 170 tgaattgaaaagggttcagagtcctgttaacctcttcttaaacctccctaaaggagtga 229
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1989 CCGTTTAGTCAAGGAAATGCTGCAGCCAAACATCTCAAAAACCTCTCTGAAGAAATTG 2048
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 230 actatcctatttcagagacttgatgagacaagctcttttcacagacatcagaaggag 289
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2049 ACTGCACGTTATATGAATACTTATTGTGGAAAAAGAAATATGAAATATTCACACAGAG 2108
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 290 acttgagcagatttgctcagga 311
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2109 ACGTTTCAGGCTTGTCATTGGA 2130
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 3
PCT-US96-12860-13
; Sequence 13, Application PC/TUS9612860
; GENERAL INFORMATION:
; APPLICANT: TULARIK, INC.
```

```
; TITLE OF INVENTION: INHIBITORS OF APOPTOSIS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/12860
; FILING DATE: 06 AUG 1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. Serial Nos. 08/512,946 & 08/569,749
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brezner, David J.
; REGISTRATION NUMBER: 24,774
; REFERENCE/DOCKET NUMBER: A-62464/DJB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)781-1989
; TELEFAX: (415)398-3249
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2862 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; PCT-US96-12860-13

Query Match 1.3%; Score 39.6; DB 5; Length 2862;
Best Local Similarity 46.9%; Pred. No. 0.15;
Matches 123; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

QY 50 tgactgttataaagcaatacacagatgacctatttgatggaatgctcgaatcggaag 109
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1869 TGACACATGTCCTTCCTGATATCTCTTGAGGCCAGTGTAATTAACAAACAGG 1928
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 110 aagtaaacatcattgtcgagaagtgagcagagatgctgctagaaggatcattaca 169
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1929 AACATGATATTATTAGACAGAAAAACACAGATACCTTACAGCAGAGAGACTTTTGACA 1988
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 170 tgaattgaaaagggttcagagtcctgttaacctcttcttaaacctccctaaaggagtga 229
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1989 CCGTTTAGTCAAGGAAATGCTGCAGCCAAACATCTTCAAAAACCTCTCTGAAGAAATTG 2048
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 230 actatcctatttcagagacttgatgagacaagctcttttcacagacatcagaaggag 289
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2049 ACTGCACGTTATATGAATACTTATTGTGGAAAAAGAAATATGAAATATTCACACAGAG 2108
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 290 acttgagcagatttgctcagga 311
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2109 ACGTTTCAGGCTTGTCATTGGA 2130
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 4
US-09-212-971-13
; Sequence 13, Application US/09212971B
; Patent No. 6107041
; GENERAL INFORMATION:
; APPLICANT: Korneluk, Robert G
; APPLICANT: Mackenzie, Alexander E
; APPLICANT: Liston, Peter
; APPLICANT: Baird, Stephen
; APPLICANT: Tsang, Benjamin K
```


SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 13
LENGTH: 3151
TYPE: DNA
ORGANISM: Mus musculus
US-09-617-053A-13

Query Match 1.3%; Score 39.6; DB 4; Length 3151;
Best Local Similarity 46.9%; Pred. No. 0.16;
Matches 123; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

OY 50 tgaactgtataaagaacacacagatgacatttggatgagatgttcgaacgcgaag 109
DB 2175 tgaacatgtcctccatcccgatgatactctcttgagccagtggaatcaaacag 2234
OY 110 aagtaaacctcttgcctcgagaaagtgagcagatgctctgaagatcattca 169
DB 2235 aacatgatatattgacacagaacacacagatcccttaacagcagagctattgaca 2294
OY 170 tgatttgaagaagggttcagagctcgttaacctcttcttaaccccttaaggagtgga 229
DB 2295 ccgttttgaatcagaggaatgtctgcagccacatcttcaaaaactctcgaaggaatg 2354
OY 230 actatccctatttcagagactgaaatgagcaaaagtccttcttcacacacacagagag 289
DB 2355 actcagcagtgatgaaaactatttggtaaaagataatgatatattccacagag 2414
OY 290 acttgagcagatttgctcagga 311
DB 2415 acgttcaggtctgcatcagga 2436

RESULT 7

PCT-US95-05922A-1
Sequence 1, Application PC/TUS9505922A

GENERAL INFORMATION:
APPLICANT: HE, ET AL.

TITLE OF INVENTION: Human Inhibitor of Apoptosis Gene 1

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,

ADDRESSEE: CECCHI, STEWART & OLSTEIN

STREET: 6 BECKER FARM ROAD

CITY: ROSELAND

STATE: NEW JERSEY

COUNTRY: USA

ZIP: 07068

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE

COMPUTER: IBM PS/2

OPERATING SYSTEM: MS-DOS

SOFTWARE: WORD PERFECT 5.1

CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/05922A

FILING DATE: 11 MAY 1995

CLASSIFICATION:
PRIOR APPLICATION DATA:

APPLICATION NUMBER:
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: FERRARO, GREGORY D.

REGISTRATION NUMBER: 36,134

REFERENCE/DOCKET NUMBER: 325800-292

TELEPHONE: 201-994-1700

TELEFAX: 201-994-1744

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1435 BASE PAIRS
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR

MOLECULE TYPE: CDNA
PCT-US95-05922A-1

Query Match 1.2%; Score 36.8; DB 5; Length 1435;
Best Local Similarity 46.2%; Pred. No. 0.69;
Matches 122; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

OY 48 aagactgtatataaagaacatcacagatgacatctatttggatgaaatgtctgacgcga 107
DB 882 ATTGACATGTGTCTCTTCCTATCTCGATATATCTTTTAAAGCCATGTAATTAATAACA 941
OY 108 agaagtaaacatcatcttgcctcgagaaagtgagcagatgctctgtagagggatca 167
DB 942 GGACATGATATATTAATAACAAAACACAGATACCTTTACAGCGAGAGAACTGATGTA 1001
OY 168 catgatttgaagaagggttcagagctcgttaacctcttcttaaatcccttaaggaag 227
DB 1002 TACCATTTTGGTTAAAGAAATGCTCGGCCAACATCTTCMAAAACTGCTTAAAGAAAT 1061
OY 228 gaactatccctatcttcagagactgagacaaagctcttcttcacagacacagaag 287
DB 1062 TGACTCTACATTGTATTAAGACTTATTGTGATGAAGATATGAACTATATTCACACGA 1121
OY 288 agacttgagcagatttgctcagga 311
DB 1122 AGATGTTTCAGGTCTGTCACACTGGA 1145

RESULT 8

US-08-569-749-1
Sequence 1, Application US/08569749

PATENT No. 6187557

GENERAL INFORMATION:
APPLICANT: Rothe, Mike

APPLICANT: Goeddel, David V

TITLE OF INVENTION: INHIBITORS OF APOPTOSIS

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/569,749

FILING DATE:
CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:
NAME: Brezner, David J.

REGISTRATION NUMBER: 24,774

REFERENCE/DOCKET NUMBER: A-62464/DJB

TELEPHONE: (415)781-1989

TELEFAX: (415)398-3249

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2589 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA

US-08-569-749-1

Query Match 1.2%; Score 36.8; DB 4; Length 2589;
Best Local Similarity 46.2%; Pred. No. 1;

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2002, 11:44:40 ; Search time 270.82 Seconds

(without alignments)
9724.921 Million cell updates/sec

Title: US-09-697-089-3

Perfect score: 3072

Sequence: 1 atgaattcataaaggacaa.....ctttaactagtaactgct 3072

Printing table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

```

1: N.Geneseq_1101.*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3072	100.0	3133	22	AA503945	Human caspase recr
2	3067.2	99.8	3545	22	AAH98254	Murine EST-derived
3	2980.6	97.0	3615	22	AA503946	Human caspase recr
4	1920.6	62.5	2950	22	AAH99581	Human protein enco
5	538.4	17.5	2735	22	AAH34171	Human colon cancer
6	242	7.9	421	22	AA114389	Probe #4322 for ge
7	242	7.9	421	22	AA135764	Probe #4450 used t
8	242	7.9	421	22	AA104213	Probe #4204 used t
9	220	7.2	220	22	AA123590	Probe #13523 for g
10	220	7.2	220	22	AA148904	Probe #17590 used
11	220	7.2	220	22	AA109206	Probe #9197 used t

12	91.8	3.0	936	22	AAF58252	Oligonucleotide D1
13	91.8	3.0	936	22	AAF58254	Oligonucleotide D1
14	91.8	3.0	936	22	AAF58257	Oligonucleotide D1
15	91.8	3.0	936	22	AAF58259	Oligonucleotide D2
16	91.8	3.0	936	22	AAF58262	Oligonucleotide D2
17	91.8	3.0	938	22	AAF58255	Oligonucleotide D1
18	91.4	3.0	936	22	AAF58252	Oligonucleotide D1
19	91.4	3.0	936	22	AAF58254	Oligonucleotide D1
20	91.4	3.0	936	22	AAF58257	Oligonucleotide D1
21	91.4	3.0	936	22	AAF58259	Oligonucleotide D2
22	91.4	3.0	936	22	AAF58262	Oligonucleotide D2
23	91.4	3.0	938	22	AAF58255	Oligonucleotide D1
24	59	1.9	5366	20	AAF58001	Gonadotropic hormo
25	59	1.9	5366	20	AA556273	Human apoptosis in
26	59	1.9	5502	17	AA130092	Neuronal apoptosis
27	59	1.9	5502	18	AA171263	Neuronal apoptosis
28	59	1.9	5984	20	AA556000	Gonadotropic hormo
29	59	1.9	5984	21	AA556272	Human apoptosis in
30	59	1.9	5984	21	AAA39808	Human Nalp CDNA.
31	59	1.9	6124	18	AA171265	Neuronal apoptosis
32	59	1.9	6133	18	AA171264	Neuronal apoptosis
33	59	1.9	6228	18	AA171266	Neuronal apoptosis
34	42	1.4	244	22	AAF58238	Oligonucleotide D1
35	41.4	1.3	4141	20	AA209251	Murine CARD-4L cDN
36	41.4	1.3	4141	22	AAF30006	Mouse CARD-4L (lon
37	39.8	1.3	3080	20	AA209248	Human CARD-4S part
38	39.8	1.3	3080	22	AAF30003	Human CARD-4S (sho
39	39.8	1.3	3382	22	AAF30002	Human CARD-4L (lon
40	39.8	1.3	4415	21	AAA95791	Apoptosis related
41	39.8	1.3	4610	22	AAH18313	Human CDNA sequenc
42	39.8	1.3	32042	20	AA209252	Human CARD-4 DNA.
43	39.8	1.3	32042	22	AAF30011	Human CARD-4 gene.
44	39.6	1.3	244	22	AAF58238	Oligonucleotide D1
45	39.6	1.3	2862	18	AA161592	Murine C-IAP. Mus

ALIGNMENTS

RESULT 1	
AA503945	
ID	AA503945 standard; cDNA; 3133 BP.
XX	
AC	AA503945;
DT	12-SEP-2001 (first entry)
XX	
DE	Human caspase recruitment domain 12 (CARD-12) cDNA.
XX	
KW	Caspase recruitment domain; CARD-12; apoptosis; stress-related pathway;
KW	cancer; viral infection; poxvirus; adenovirus; autoimmune disorder;
KW	systemic lupus erythematosus; arthritis; neurological disorder; stroke;
KW	Alzheimer's disease; amyotrophic lateral sclerosis; haematologic disease;
KW	aplastic anaemia; myocardial infarction; inflammatory disorder;
KW	Crohn's disease; insulin-dependent diabetes; contact dermatitis;
KW	psoriasis; graft rejection; bacterial infection; lepromatous leprosy;
KW	tuberculosis; ischaemic brain injury; hypoxic brain injury; ss;
KW	kidney ischaemia; reperfusion injury; acute bacterial meningitis;
KW	excitotoxic brain damage; liver disease.
OS	Homo sapiens.
XX	
EH	Key
FT	Location/Qualifiers
FT	36..3110
FT	/*tag= a
FT	/product= "Human CARD-12"
PN	WO200130971-A2.
XX	
PD	03-MAY-2001.
XX	
PF	26-OCT-2000; 2000WO-US29643.
XX	

OY 1561 cttccatcgccaagagcctctctgagacagaatcttgcagaagtgcataaacacc 1620
|||||
Db 1596 cttccatcgccaagagcctctctgagacagaatcttgcagaagtgcataaacacc 1655
OY 1621 actgagcagaatctctgaaagccataacatcaatcccttctgtagagtgcacat 1680
Db 1656 actgagcagaatctctgaaagccataacatcaatcccttctgtagagtgcacat 1715
OY 1681 ttatctcagaagagtagctcccaatcaagccctgagccaagaatttggaagcttcttcaa 1740
Db 1716 ttatctcagaagagtagctcccaatcaagccctgagccaagaatttggaagcttcttcaa 1775
OY 1741 ggtcaaaagcttatataataactcaagcgaaccccgatlaacttaattgacttcttga 1800
Db 1776 ggtcaaaagcttatataataactcaagcgaaccccgatlaacttaattgacttcttga 1835
OY 1801 catttgcccaattctgcaagctgctctggaactcaatgaacttgaagggagct 1860
Db 1836 catttgcccaattctgcaagctgctctggaactcaatgaacttgaagggagct 1895
OY 1861 atggtctatagggaagagcctgagcaagacacagctggaatccacatggaagagccca 1920
Db 1896 atggtctatagggaagagcctgagcaagacacagctggaatccacatggaagagccca 1955
OY 1921 gaaacctacatcccaagcagcgtatcttcttctcaactggaagcagaatccag 1980
Db 1956 gaaacctacatcccaagcagcgtatcttcttctcaactggaagcagaatccag 2015
OY 1981 actctgagagtcacacctccggagattccagcaagtgaataagcaagatacacatc 2040
Db 2016 actctgagagtcacacctccggagattccagcaagtgaataagcaagatacacatc 2075
OY 2041 gggaaaaatattcagctgcacacagcctcagcgtcgaataataagaagtgcgtgtg 2100
Db 2076 gggaaaaatattcagctgcacacagcctcagcgtcgaataataagaagtgcgtgtg 2135
OY 2101 gctggaagcctcagcttctgctccagcaccctgtaagaacattatctctcagtgtaa 2160
Db 2136 gctggaagcctcagcttctgctccagcaccctgtaagaacattatctctcagtgtaa 2195
OY 2161 gccagtcctccacacatagaatgagcagacatcacatctgtacaacacctgtaaac 2220
Db 2196 gccagtcctccacacatagaatgagcagacatcacatctgtacaacacctgtaaac 2255
OY 2221 ttgagtatctatgactacagaatcaacgctgcgcgggtgtctactgcagacgcttgg 2280
Db 2256 ttgagtatctatgactacagaatcaacgctgcgcgggtgtctactgcagacgcttgg 2315
OY 2281 aacttgaagacactatacaagctcataatgatacaataagaatgaagaagatgct 2340
Db 2316 aacttgaagacactatacaagctcataatgatacaataagaatgaagaagatgct 2375
OY 2341 ataaaaactagctgaagcctgtaaaaaacctgtaagaagtgttlaattcatlttgaccac 2400
Db 2376 ataaaaactagctgaagcctgtaaaaaacctgtaagaagtgttlaattcatlttgaccac 2435
OY 2401 ttgtctgacatctgagagaggaatgataatcaatgataagctctctgtaagaagccctgt 2460
Db 2436 ttgtctgacatctgagagaggaatgataatcaatgataagctctctgtaagaagccctgt 2495
OY 2461 gaccttgaagaatcaatgaatgctcctgctgctgtctgtaagaagcagtgaaacctcta 2520
Db 2496 gaccttgaagaatcaatgaatgctcctgctgctgtctgtaagaagcagtgaaacctcta 2555
OY 2521 gctcagaatcttcaaatctgtgtaaacctgagcattctgtatctacagaagaattaccgt 2580
Db 2556 gctcagaatcttcaaatctgtgtaaacctgagcattctgtatctacagaagaattaccgt 2615
OY 2581 gaaaaaagtgaaatgaaagcctctcatgaactgatacgcaagatgaagcgtctgtaaacag 2640
Db 2616 gaaaaaagtgaaatgaaagcctctcatgaactgatacgcaagatgaagcgtctgtaaacag 2675
OY 2641 ctcaacgcactgactgctcctggtggtgtgactgtaagcagacgctgagagcctgtg 2700

Db 2676 ctcacacgcactgactgctcctggtggtgtgacgctgcaagcagcgtgagcgtgtg 2735
OY 2701 aaacattgagagaggtccacacactcgtcaagcttgggttgtaaaaaactgagactcaca 2760
Db 2736 aaacattgagagaggtccacacactcgtcaagcttgggttgtaaaaaactgagactcaca 2795
OY 2761 gatacagaagattagaattttgagtgatcttttttgtaaaagaacccctgtaaaactccaag 2820
Db 2796 gatacagaagattagaattttgagtgatcttttttgtaaaagaacccctgtaaaactccaag 2855
OY 2821 cagttgaatttgccgggaatcgtgtgagcaagtgaatgagtgagtgagtgagtgagtgta 2880
Db 2856 cagttgaatttgccgggaatcgtgtgagcaagtgaatgagtgagtgagtgagtgagtgta 2915
OY 2881 ttggaacatcttaagaacattagtgcttttttgacttagtactaaagaatttactctgat 2940
Db 2916 ttggaacatcttaagaacattagtgcttttttgacttagtactaaagaatttactctgat 2975
OY 2941 ccagcattagtcagaagaacttagccaagtgtatcccaagttactttctgcaagaagct 3000
Db 2976 ccagcattagtcagaagaacttagccaagtgtatcccaagttactttctgcaagaagct 3035
OY 3001 aggcctgtgtggtgcaatttgatgatgatactccagtggtattacagtggtctttaaa 3060
Db 3036 aggcctgtgtggtgcaatttgatgatgatactccagtggtattacagtggtctttaaa 3095
OY 3061 ctgataactgct 3072
Db 3096 ctgataactgct 3107

RESULT 2
AAH98254
ID AAH98254 standard; cDNA: 3545 BP.
XX
AC AAH98254;
XX
DT 12-OCT-2001 (first entry)
XX
DE Murine EST-derived coding sequence SPO ID NO: 111.
XX
KW Human: sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder;
KW biodiversity; gene therapy; nutrition; ss.
OS Mus musculus.
XX
PN W0200154477-A2.
XX
PD 02-AUG-2001.
XX
PF 25-JAN-2001; 2001WO-US02687.
XX
PR 25-JAN-2000; 2000US-0491404.
PR 17-JUL-2000; 2000US-0617746.
PR 03-AUG-2000; 2000US-0631451.
PR 15-SEP-2000; 2000US-0663870.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V,
PI Cao Y, Drmanac RA, Zhang J, Werhman T;
XX
DR WPI: 2001-476164/51.
DR P-PSDB; AAM23595.
XX
PT Isolated polypeptide for treatment of diseases; diagnostics; raising
PT antibodies and research use -
XX
PS Claim 1; Page 250-251; 1275bp; English.
XX

CC The present invention provides the protein and coding sequences of novel
 CC proteins from a variety of organisms, including human, dog, cat, horse,
 CC cow, pig, hamster, monkey, macaque, yeast, bacterium, fruit fly, sea
 CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
 CC from the organism of interest. They can be used in diagnostics,
 CC forensics, gene mapping, identification of mutations, to assess
 CC biodiversity and for nutritional purposes. The present sequence is a cDNA
 CC of the invention.

XX Sequence 3545 BP; 1038 A; 755 C; 816 G; 936 T; 0 other;

Query Match 99.8%; Score 3067.2; DB 22; Length 3545;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 3069; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 atgaattcataaaggacatagcagccctattcaagaatggagatgtatata 60
 DB |||||||
 DB 232 atgaattcataaaggacatagcagccctattcaagaatggagatgtatata 231
 OY 61 aagcaaatcacagatgacctattgtatgaatgtctgacgaaagaataaacatc 120
 DB |||||||
 DB 292 aagcaaatcacagatgacctattgtatgaatgtctgacgaaagaataaacatc 351
 OY 121 atttgtcgcgagaaggtgagcagaagatgctgtagagagatcatcatatgtattgaaa 180
 DB |||||||
 DB 352 atttgtcgcgagaaggtgagcagaagatgctgtagagagatcatcatatgtattgaaa 411
 OY 181 aagggtctcagagctcctgtacacctcttcttaaaccttaagaagatggagactctctca 240
 DB |||||||
 DB 412 aagggtctcagagctcctgtacacctcttcttaaaccttaagaagatggagactctctca 471
 OY 241 ttccagacttgaaatggacaaagtcttcttcacacagacatcagaagaagacttgcagat 300
 DB |||||||
 DB 472 ttccagacttgaaatggacaaagtcttcttcacacagacatcagaagaagacttgcagat 531
 OY 301 ttggtctcagagatgttaagaagacttgcacatccccactcttcttgactttatccccct 360
 DB |||||||
 DB 532 ttggtctcagagatgttaagaagacttgcacatccccactcttcttgactttatccccct 591
 OY 361 ggtgaagatattgacatcttcttaacttgaaaagacaccttcacagaacctgcctgtg 420
 DB |||||||
 DB 592 ggtgaagatattgacatcttcttaacttgaaaagacaccttcacagaacctgcctgtg 651
 OY 421 aggaagagacaaacacatcacgcgtgagacagctgacacctgaaatgacctctgcagct 480
 DB |||||||
 DB 652 aggaagagacaaacacatcacgcgtgagacagctgacacctgaaatgacctctgcagct 711
 OY 481 ctccagagccctctgacatctatgaaggagatctgcgcaaggacagctcctgcctgcag 540
 DB |||||||
 DB 712 ctccagagccctctgacatctatgaaggagatctgcgcaaggacagctcctgcctgcag 771
 OY 541 cgcattgcacatgctctgcggcctccggaagaagtgcgaagctctgcacgaatccaatctgc 600
 DB |||||||
 DB 772 cgcattgcacatgctctgcggcctccggaagaagtgcgaagctctgcacgaatccaatctgc 831
 OY 601 ttctctctcgcgtctcaagcagggcccgagggtggacttttgaaacctctgtgtacatc 660
 DB |||||||
 DB 832 ttctctctcgcgtctcaagcagggcccgagggtggacttttgaaacctctgtgtacatc 891
 OY 661 ctgagatactcctgcgcaaatcagaagaagacatcatatgacatgctgtctgaaagctgcgg 720
 DB |||||||
 DB 892 ctgagatactcctgcgcaaatcagaagaagacatcatatgacatgctgtctgaaagctgcgg 951
 OY 721 cagaaggtctcttctctcttgatgctacaatgaattcaaaccccaagacatgcccagaa 780
 DB |||||||
 DB 952 cagaaggtctcttctctcttgatgctacaatgaattcaaaccccaagacatgcccagaa 1011
 OY 781 atcgaagccctgtataaagaaacacacgccttcaagaacatgtctatctaccactaac 840
 DB |||||||
 DB 1012 atcgaagccctgtataaagaaacacacgccttcaagaacatgtctatctaccactaac 1071
 OY 841 actgaagctcctgagacatacagcagatttgttgcctctgactgtcgtgaggtgggatactg 900

DB |||||||
 DB 1072 actgaagctcctgagacatacagcagatttgttgcctctgactctgtcaggtggggatactg 1131
 OY 901 aacgaatacagcccccaggtctctcatccaggaagtgtctgaaggagctgtcgtgaagc 960
 DB |||||||
 DB 1132 aacgaatacagcccccaggtctctcatccaggaagtgtctgaaggagctgtcgtgaagc 1191
 OY 961 ttgttctccaaattccagaatcccaaggtgtcttgagaaatctcatgaagaccctctctt 1020
 DB |||||||
 DB 1192 ttgttctccaaattccagaatcccaaggtgtcttgagaaatctcatgaagaccctctctt 1251
 OY 1021 gtgttctcactgtgtcaatccagatggtgtgaagatgttccactctcacacaaca 1080
 DB |||||||
 DB 1252 gtgttctcactgtgtcaatccagatggtgtgaagatgttccactctcacacaaca 1311
 OY 1081 acgctgtccatccctctctatgtatctgtatcaagaanaacaacaacaataaagct 1140
 DB |||||||
 DB 1312 acgctgtccatccctctctatgtatctgtatcaagaanaacaacaacaataaagct 1371
 OY 1141 gtgtctgcaagtgactctatctcgagagcttgaccactgttgagactagctctggaagt 1200
 DB |||||||
 DB 1372 gtgtctgcaagtgactctatctcgagagcttgaccactgttgagactagctctggaagt 1431
 OY 1201 gtgttctcccaagtttgatttccgaactgcagagatgtgtccagcgtgtaatgagatgtc 1260
 DB |||||||
 DB 1432 gtgttctcccaagtttgatttccgaactgcagagatgtgtccagcgtgtaatgagatgtc 1491
 OY 1261 ctgtcgaacacttggtccctctgtaaaatatacagcttcaagaaggttcaagccaaagtataa 1320
 DB |||||||
 DB 1492 ctgtcgaacacttggtccctctgtaaaatatacagcttcaagaaggttcaagccaaagtataa 1551
 OY 1321 ttcttccaaagtcatctcagagatcacacagcaggaagactcagcagttattgacg 1380
 DB |||||||
 DB 1552 ttcttccaaagtcatctcagagatcacacagcaggaagactcagcagttattgacg 1611
 OY 1381 ttctcatgagccagagaggtgagccaaagggaatggttacttgcagaanaatggttccatt 1440
 DB |||||||
 DB 1612 ttctcatgagccagagaggtgagccaaagggaatggttacttgcagaanaatggttccatt 1671
 OY 1441 tccgacatcatccactcttaagcagcctctccgagacacctgtgtgtatcctgtgaa 1500
 DB |||||||
 DB 1672 tccgacatcatccactcttaagcagcctctccgagacacctgtgtgtatcctgtgaa 1731
 OY 1501 gccacaaaggctgttatgaagacacctgcagacagtgatatacaacaggtcctctcga 1560
 DB |||||||
 DB 1732 gccacaaaggctgttatgaagacacctgcagacagtgatatacaacaggtcctctcga 1791
 OY 1561 ctctcatcgcgaagaggtcctctctgagacaggaatcttgcagaagtgtgaaaaaacac 1620
 DB |||||||
 DB 1792 ctctcatcgcgaagaggtcctctctgagacaggaatcttgcagaagtgtgaaaaaacac 1851
 OY 1621 actgagcaagaattctgaaagccataaacaatccttctgaaagtctgacatcat 1680
 DB |||||||
 DB 1852 actgagcaagaattctgaaagccataaacaatccttctgaaagtctgacatcat 1911
 OY 1681 ttatatacagaagatgataatccaaatcagccctgagccaaagatttgaagcttcttcaa 1740
 DB |||||||
 DB 1912 ttatatacagaagatgataatccaaatcagccctgagccaaagatttgaagcttcttcaa 1971
 OY 1741 ggtataaagcttatataatcaactcagaggaacatacccgatattcttgaactctttaa 1800
 DB |||||||
 DB 1972 ggtataaagcttatataatcaactcagaggaacatacccgatattcttgaactctttaa 2031
 OY 1801 catttgcacaaatgtgcaagtgtctggaacttattaaacttgacattttagggagact 1860
 DB |||||||
 DB 2032 catttgcacaaatgtgcaagtgtctggaacttattaaacttgacattttagggagact 2091
 OY 1861 atgtgtctcatggtgaaagagctgcagagaagacacaggttgaatccacatggaagggccca 1920
 DB |||||||
 DB 2092 atgtgtctcatggtgaaagagctgcagagaagacacaggttgaatccacatggaagggccca 2151
 OY 1921 gaaactacatcccaagcaggtctgtatcttcttctcaactcgtgaagcagaatcaggt 1980
 DB |||||||

Db 2152 gaacatcattccacagcaggcgtglatcttcttctcaactggaagcaggaattcagg 2211
 Qy 1981 actctggaggtcaacactccgggatttcagcaagtgtgaataagcaatatacatctg 2040
 Db 2212 actctggaggtcaacactccgggatttcagcaagtgtgaataagcaatatacatctg 2271
 Qy 2041 gggaaaatatcagctcgcacaaagcctcaggtcgtcaataaagaatgtgctgtgtg 2100
 Db 2272 gggaaaatatcagctcgcacaaagcctcaggtcgtcaataaagaatgtgctgtgtg 2331
 Qy 2101 gctggaaagcctcagttgttctcctcagcaccctgttaagaacattatctcctcag 2160
 Db 2332 gctggaaagcctcagttgttctcctcagcaccctgttaagaacattatctcctcag 2391
 Qy 2161 gccagtcctccaccatagaagatgagagacatcattctgttaacaaacttgaacc 2220
 2392 gccagtcctccaccatagaagatgagagacatcattctgttaacaaacttgaacc 2451
 Qy 2221 ttgagttatcagacctcaagaatcaacgctgcgggtgtgtcagctcagcagcttggt 2280
 Db 2452 ttgagttatcagacctcaagaatcaacgctgcgggtgtgtcagctcagcagcttggt 2511
 Qy 2281 aacttgaagaacctacaaagctcataatgtgataacataaagaatgtgaagaagatgct 2340
 Db 2512 aacttgaagaacctacaaagctcataatgtgataacataaagaatgtgaagaagatgct 2571
 Qy 2341 ataataactagctgaagcctgtgaaaaaacttgaagaagatgtgtttatctcatttgccac 2400
 Db 2572 ataataactagctgaagcctgtgaaaaaacttgaagaagatgtgtttatctcatttgccac 2631
 Qy 2401 ttgctcagcatctggaggaagatgataacataagctcctcagtggaaccctgt 2460
 Db 2632 ttgctcagcatctggaggaagatgataacataagctcctcagtggaaccctgt 2691
 Qy 2461 gaccttgaagaatcaactagctcctcctgctgtgtctgtcgaatgtgacgtgaaatccta 2520
 Db 2692 gaccttgaagaatcaactagctcctcctgctgtgtctgtcgaatgtgacgtgaaatccta 2751
 Qy 2521 gctcagaatcttcacaaatttggtcaaaactgagcatcttctgtatttcacgaataatctg 2580
 Db 2752 gctcagaatcttcacaaatttggtcaaaactgagcatcttctgtatttcacgaataatctg 2811
 Qy 2581 gaaaaaagtgaagaatgagcctctcattgactgacgacgagatgaagctgtcagaacag 2640
 Db 2812 gaaaaaagtgaagaatgagcctctcattgactgacgacgagatgaagctgtcagaacag 2871
 Qy 2641 ctcaacgcagctgagctgctcctggtggtgtgacgtgtgcaagcagcctgagcagcctgtg 2700
 Db 2872 ctcaacgcagctgagctgctcctggtggtgtgacgtgtgcaagcagcctgagcagcctgtg 2931
 Qy 2701 aaacatttggggaggtgtcccaactcgtcaagcttggtgttgaaaaaacttgagactcaca 2760
 Db 2932 aaacatttggggaggtgtcccaactcgtcaagcttggtgttgaaaaaacttgagactcaca 2991
 Qy 2761 gatacagagatagaattttagtgatcttcttgaagaagaacctctgaanaactccag 2820
 Db 2992 gatacagagatagaattttagtgatcttcttgaagaagaacctctgaanaactccag 3051
 Qy 2821 cagttgaatttggcgggaacatcgtgtgacgctgtgataagctgtgctcctcacaagggtga 2880
 Db 3052 cagttgaatttggcgggaacatcgtgtgacgctgtgataagctgtgctcctcacaagggtga 3111
 Qy 2881 tttaggaatcttaagcaattttagtcttcttgaactttagtactaaagaatttctactatg 2940
 Db 3112 tttaggaatcttaagcaattttagtcttcttgaactttagtactaaagaatttctactatg 3171
 Qy 2941 ccaagatattgacagaanaacttagccaagtgtatccaagttaacttctcgtgaagaact 3000
 Db 3172 ccaagatattgacagaanaacttagccaagtgtatccaagttaacttctcgtgaagaact 3231
 Qy 3001 aggcctgttgggtggcaatttgaatgaatgaatcagctcagtgatatacaggtgtctttaa 3060
 Db 3232 aggcctgttgggtggcaatttgaatgaatgaatcagctcagtgatatacaggtgtctttaa 3291

Qy 3061 ctgagtaactgct 3072
 Db 3292 ctgagtaactgct 3303

RESULT 3
 AAS03946
 ID AAS03946 standard; DNA; 3615 BP.
 XX
 AC AAS03946;
 XX
 DT 12-SEP-2001 (first entry)
 XX
 DE Human caspase recruitment domain 12 (CARD-12) genomic DNA.
 XX
 KW Caspase recruitment domain; CARD-12; apoptosis; stress-related pathway;
 KW cancer; viral infection; poxvirus; adenovirus; autoimmune disorder;
 KW systemic lupus erythematosus; arthritis; neurological disorder; stroke;
 KW Alzheimer's disease; amyotrophic lateral sclerosis; hematologic disease;
 KW aplastic anaemia; myocardial infarction; inflammatory disorder;
 KW Crohn's disease; insulin-dependent diabetes; contact dermatitis;
 KW psoriasis; graft rejection; bacterial infection; lepromatous leprosy;
 KW tuberculosis; ischaemic brain injury; hypoxic brain injury; ds;
 KW kidney ischaemia; reperfusion injury; acute bacterial meningitis;
 KW excitotoxic brain damage; liver disease.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..3615
 FT /tag= a
 FT /product= "Human CARD-12"
 FT
 PN WO200130971-A2.
 XX
 PD 03-MAY-2001.
 XX
 PF 26-OCT-2000; 2000WO-US29643.
 XX
 PR 27-OCT-1999; 99US-0161822.
 XX
 PA (MILL-) MILLENNIUM PHARM INC.
 XX
 PI Bertin J, Robison KE;
 DR WPI, 2001-308628/32.
 DR P-PSDB; AAU02881.
 XX
 PT Isolated caspase recruitment domain-12 polypeptide and nucleic acids
 PT encoding them, useful for treating and diagnosing disorders associated
 PT with abnormal apoptosis such as cancer, arthritis and Alzheimer's
 PT disease -
 XX
 PS Disclosure; Fig 2; 93pp; English.
 XX

The sequence represents a genomic DNA which encodes the human caspase recruitment domain 12 (CARD-12) polypeptide. CARD domains are found in a number of proteins that transmit signals that activate apoptosis and inflammatory pathways in response to stress and other stimuli. Therefore, CARD-12 and its corresponding nucleic acid may be used in treatment and diagnosis of patients suffering from disorders associated with an abnormal level (an increase or a decrease) of apoptotic cell death or abnormal activity of stress-related pathways. The disorders include cancer, viral infections (e.g. caused by poxviruses, adenoviruses), autoimmune disorders (e.g. systemic lupus erythematosus, arthritis), neurological disorders (e.g. Alzheimer's disease, amyotrophic lateral sclerosis), hematologic diseases (e.g. aplastic anaemia, myocardial infarction, stroke), inflammatory and immune system disorders (e.g. Crohn's disease, insulin-dependent diabetes, contact dermatitis, psoriasis, graft rejection), bacterial infections (e.g. tuberculosis, lepromatous leprosy), ischaemic and hypoxic brain injury, kidney ischaemia/reperfusion injury, excitotoxic brain damage, acute bacterial

CC meningitis and liver disease.
XX
SQ Sequence 3615 BP; 1041 A; 811 C; 845 G; 918 T; 0 other;

Query Match 97.0%; Score 2980.6; DB 22; Length 3615;
Best Local Similarity 98.1%; Pred. No. 0;
Matches 3047; Conservative 0; Mismatches 4; Indels 54; Gaps 1;

QY 2 tgaattcataaagacataagccagcagcccttattcaagaatgaggatgactgtataa 61
DB 419 tgaattcataaagacataagccagcagcccttattcaagaatgaggatgactgtataa 478
QY 62 agcaaatcacagatgacatttgtatgaaatgttctgaaatcgagaaatgataca 121
DB 479 agcaaatcacagatgacatttgtatgaaatgttctgaaatcgagaaatgataca 538
QY 122 ttgtcgcgagaagatgagcgaggaatgctgtatgaggaatcattacatgatttga 161
DB 539 ttgtcgcgagaagatgagcgaggaatgctgtatgaggaatcattacatgatttga 598
QY 182 aggtgtcagagtcctgttaacctcttcttaaatcccttaagaggtgaaatccttat 241
DB 599 aggtgtcagagtcctgttaacctcttcttaaatcccttaagaggtgaaatccttat 658
QY 242 ttcaggaacttgatggacaaa----- 262
DB 659 ttcaggaacttgatggacaaaagtttgagagacacagaattggtctcttcaacata 718
QY 263 -----gtcttttcatcagacatcagagaagagacttggagatttgctc 307
DB 719 cctctctctaataagctcttcttcaacacagacagaagagacttggagatttgctc 778
QY 308 aggatttaaggagactgttaccatacccatcttcttgaacttlaacccctgtgaa 367
DB 779 aggatttaaggagactgttaccatacccatcttcttgaacttlaacccctgtgaa 838
QY 368 atattgacattattttaaacttgaaaaagcacttcaagaacctgtccgttggagaa 427
DB 839 atattgacattattttaaacttgaaaaagcacttcaagaacctgtccgttggagaa 898
QY 428 accaaacacatcacgcgtgtgagcagctgaccttgaatggcccttcgcaagccttaca 467
DB 899 accaaacacatcacgcgtgtgagcagctgaccttgaatggcccttcgcaagccttaca 958
QY 488 gccctgtacatctgaaggaggaaatctgycgaaagcgaagtcacactgtctgacagcat 547
DB 959 gccctgtacatctgaaggaggaaatctgycgaaagcgaagtcacactgtctgacagcat 1018
QY 548 ccattgtctcggggctccggaagaatgcaaggctctgaccaaagtccaattcgtctcttc 607
DB 1019 ccattgtctcggggctccggaagaatgcaaggctctgaccaaagtccaattcgtctcttc 1078
QY 608 tccgtctcagcagggcccgaggtgtgagcttcttgaaaacctctgttgaactcttgata 667
DB 1079 tccgtctcagcagggcccgaggtgtgagcttcttgaaaacctctgttgaactcttgata 1138
QY 668 taacctgtgacaatlcagaaagcagacatlcattggccaatgctgtcgtgaagctgagcgaag 727
DB 1139 taacctgtgacaatlcagaaagcagacatlcattggccaatgctgtcgtgaagctgagcgaag 1198
QY 728 tctcttctctcttctgtatgactacaaatgaaatlcgaagcccgagaacttgcagaaatcgaag 1258
DB 1199 tctcttctctcttctgtatgactacaaatgaaatlcgaagcccgagaacttgcagaaatcgaag 1258
QY 788 ccctgtataaagaaaccccgcttcaagaacatggttcatcgtccaccactacacactgagt 847
DB 1259 ccctgtataaagaaaccccgcttcaagaacatggttcatcgtccaccactacacactgagt 1318
QY 848 gccctgagacacatagcgagtttgtgtcccttgactgtgaggtgagggaatgacagaag 907
DB 1319 gccctgagacacatagcgagtttgtgtcccttgactgtgaggtgagggaatgacagaag 1378

QY 908 acagcgcccaagctctcatccgagaagtgctgatacaaggagctgtcgaagctgtgtgc 967
DB 1379 acagcgcccaagctctcatccgagaagtgctgatacaaggagctgtcgaagctgtgtgc 1438
QY 968 tccaattcagaataatccagtgcttggaggaatctcatgaaagaccctctcttgggtca 1027
DB 1439 tccaattcagaataatccagtgcttggaggaatctcatgaaagaccctctcttgggtca 1498
QY 1028 tcaactgtgcaatccagatggtgtgaaagtgtgtccactccacacacaaacacgcgtc 1087
DB 1499 tcaactgtgcaatccagatggtgtgaaagtgtgtccactccacacacaaacacgcgtc 1558
QY 1088 tcaatacctctatgatactgttgaatacagaanaaacacacacataaaggtgtgtgc 1147
DB 1559 tcaatacctctatgatactgttgaatacagaanaaacacacacataaaggtgtgtgc 1618
QY 1148 caagtgtctcatctcggaagctgtgacacactgtgtgaaagcctgactgtgaaaggtgtgtct 1207
DB 1619 caagtgtctcatctcggaagctgtgacacactgtgtgaaagcctgactgtgaaaggtgtgtct 1678
QY 1208 cccaagtttgatttctgaactgagatgtgtccagcgtgaaatgagatgtccctgtga 1267
DB 1679 cccaagtttgatttctgaactgagatgtgtccagcgtgaaatgagatgtccctgtga 1738
QY 1268 caactggtcctcctgtgaaatatacagctcaaaaggttcaagccaaagtataaattcttc 1327
DB 1739 caactggtcctcctgtgaaatatacagctcaaaaggttcaagccaaagtataaattcttc 1798
QY 1328 acaagttatcccgaggatatacagcagaagacagagactcagagtttttgaagttatg 1367
DB 1799 acaagttatcccgaggatatacagcagaagacagagactcagagtttttgaagttatg 1858
QY 1388 agccagagaggtgtgaccaaagggaatgttactgtgacagaatgtgttcatctcgaca 1447
DB 1859 agccagagaggtgtgaccaaagggaatgttactgtgacagaatgtgttcatctcgaca 1918
QY 1448 ttacatccacttatagcagctcctccgttaccactgtgtggatcctgtgtgaagcacca 1507
DB 1919 ttacatccacttatagcagctcctccgttaccactgtgtggatcctgtgtgaagcacca 1978
QY 1508 gggctgttataagagcacctcgcagcagtggtatacaacacggctcgtccttcggaattcca 1567
DB 1979 gggctgttataagagcacctcgcagcagtggtatacaacacggctcgtccttcggaattcca 2038
QY 1568 tcgcgaagagcctcctctgtgagacagaagatcttgcagaagtgtgtgaaacacacactgagc 1627
DB 2039 tcgcgaagagcctcctctgtgagacagaagatcttgcagaagtgtgtgaaacacacactgagc 2098
QY 1628 aagaatcttgaaagcacaataacataatctcctttagaggtgtgcatcattatatac 1687
DB 2099 aagaatcttgaaagcacaataacataatctcctttagaggtgtgcatcattatatac 2158
QY 1688 aagaagttataatccaatccagcctcgagcagaagaatttgaagcttcttccaaggtataa 1747
DB 2159 aagaagttataatccaatccagcctcgagcagaagaatttgaagcttcttccaaggtataa 2218
QY 1748 gcttataatcaactcagaggaacatacccgattactatttgaacttcttgaacatttgc 1807
DB 2219 gcttataatcaactcagaggaacatacccgattactatttgaacttcttgaacatttgc 2278
QY 1808 ccaattgtcgaagtgtcctgtgaacttcaactgtgaacttltatggggagctatgctc 1867
DB 2279 ccaattgtcgaagtgtcctgtgaacttcaactgtgaacttltatggggagctatgctc 2338
QY 1868 catgaggaaagcgtgcagaaagacaaagtgtgaatccacatggaagagcccgaaacct 1927
DB 2339 catgaggaaagcgtgcagaaagacaaagtgtgaatccacatggaagagcccgaaacct 2398
QY 1928 acattccagcagaggtgtatcttcttcaacttgaagcaggaatttcaaggaacttgc 1987
DB 2399 acattccagcagaggtgtatcttcttcaacttgaagcaggaatttcaaggaacttgc 2458
QY 1988 aggtcaactcgggattttagcagaagttgaaataagcagaatatcacatctgtgggaaa 2047

|||||
2459 agtcaactccgagattcagcaagtgtgaataagcaagatacatcagatactg99gaaaa 2518
OY 2048 tattcaactcgcacaaagcctcagctgcgtcaataaagaagtgtgctgtgtgaa 2107
DB 2519 tattcaactcgcacaaagcctcagctgcgtcaataaagaagtgtgctgtgaa 2578
OY 2108 gctcaagtttgtcctcagcagcctgtgaagaacattatctcctcatgtgtgaagccaa 2167
DB 2579 gccacagtttgtctcctcagcagcctgtgaagaacattatctcctcatgtgtgaagccaa 2638
OY 2168 cctcaccatagaagaatgtagagagacacatctgttaacaaacctgaaacctgtgaa 2227
DB 2639 cctcaccatagaagaatgtagagagacacatctgttaacaaacctgaaacctgtgaa 2698
2228 ttctagacctacagaatacagagctgcgagtggtgtgacgtgacgtgtgtgaactgtga 2287
2699 ttctagacctacagaatacagagctgcgagtggtgtgacgtgacgtgtgtgaactgtga 2738
OY 2288 agaaccctacaagaatcgtatgaatacagaatgaagaagaatgtatlaaac 2347
DB 2759 agaaccctacaagaatcgtatgaatacagaatgaagaagaatgtatlaaac 2818
OY 2348 tagctgaagcctgtaaaaaacctgaagaagatgtgttatattcatgtgaccactgtgtc 2407
DB 2819 tagctgaagcctgtaaaaaacctgaagaagatgtgttatattcatgtgaccactgtgtc 2878
OY 2408 acattgagagagggaatgattacatagtaagaatgtctgtcaagtgaaacctgtacctgt 2467
DB 2879 acattgagagagggaatgattacatagtaagaatgtctgtcaagtgaaacctgtacctgt 2938
OY 2468 aagaatacctaattgctcctcgtcgtctgtctgtcaaatgtcaagtgaaacctgtaccta 2527
DB 2939 aagaatacctaattgctcctcgtcgtcgtctgtcaaatgtcaagtgaaacctgtaccta 2998
OY 2528 atcttcaaatgtgttcaaaaacctgaagacatctgtatctatcagaanaatcacctgtgaa 2587
DB 2999 atcttcaaatgtgttcaaaaacctgaagacatctgtatctatcagaanaatcacctgtgaa 3058
OY 2588 atggaataatgaagcctctcctcgtcgtcgtcgtcaagtgaaacctgtgaaacctgtgaa 2647
DB 3059 atggaataatgaagcctctcctcgtcgtcgtcgtcaagtgaaacctgtgaaacctgtgaa 3118
OY 2648 caccgatgtcgtcctgtgagcgtgtacgtgtcaagcagcctgtgacagcctgtgtgaacatt 2707
DB 3119 caccgatgtcgtcctgtgagcgtgtgtacgtgtcaagcagcctgtgacagcctgtgtgaacatt 3178
OY 2708 tgaagagagtcctcacaactcgtlcaagctgtggtgtgaaanaactgtgagactcagatacag 2767
DB 3179 tgaagagagtcctcacaactcgtlcaagctgtggtgtgaaanaactgtgagactcagatacag 3238
OY 2768 agatagaatatttggtgtgcatcttttggaagaacccctgtgaaanaactcagacagctgtga 2827
DB 3239 agatagaatatttggtgtgcatcttttggaagaacccctgtgaaanaactcagacagctgtga 3298
OY 2828 attgagcgagaaatcgtgtgagcagtgatgtgagcgtgtgctgtcctcatgtgtatttgaga 2887
DB 3299 attgagcgagaaatcgtgtgagcagtgatgtgagcgtgtgctgtcctcatgtgtatttgaga 3358
OY 2888 atcttaagcaatgtgtgtttttgacttagtactaaagaattctacatcagatcagacat 2947
DB 3359 atcttaagcaatgtgtgtttttgacttagtactaaagaattctacatcagatcagacat 3418
OY 2948 tagtcagagaanaacttagcagaatgttatccagaatgaacttctgtcagaagaactgtgactgt 3007
DB 3419 tagtcagagaanaacttagcagaatgttatccagaatgaacttctgtcagaagaactgtgactgt 3478
OY 3008 ttgggtgtgcaattgtgatgtatgatctcaagtgttatcatcagagctgt 3052
DB 3479 ttgggtgtgcaattgtgatgtatgatctcaagtgttatcatcagagctgt 3523

RESULT 4

AAH9581
ID AAH9581 standard; cDNA; 2950 BP.
XX
AC AAH9581;
XX
DT 16-OCT-2001 (first entry)
XX
DE Human protein encoding cDNA sequence SEQ ID NO:416.
XX
KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
KW antiinflammatory; antirheumatic; antirheumatic; immunosuppressive;
KW antibacterial; endocrine; cardiac; central nervous system; virucide;
KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
KW antiaggregant; haemostatic; vulnery; antileuk; osteopathic; eczema;
KW dermatological; antiallergic; antidiabetic; antidiabetic; cytostatic;
KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
KW thrombocytopenia; osteoporosis; severe combined immunodeficiency;
KW allergic rhinitis; diabetes; multiple sclerosis; depression;
KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
KW neurological disorder; ss.
OS Homo sapiens.
XX
PN WO200153455-A2.
XX
PD 26-JUL-2001.
XX
PF 22-DEC-2000; 2000WO-US35017.
XX
PR 23-DEC-1999; 99US-0471275.
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
XX
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI: 2001-457603/49.
DR P-PSDB: AAM25640.
XX
PT Isolated human polynucleotides encoding polypeptides, useful for the
PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
XX
PS Claim 1: Page 511-512; 1217pp; English.
XX
CC AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
CC AAM25963. The proteins can have activities based on the tissues and
CC cells they are expressed in, such as: antinflammatory; antirheumatic;
CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac;
CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;
CC cardiovascular; antianaemic; antiaggregant; haemostatic; vulnery;
CC antileuk; osteopathic; dermatological; antiallergic; antidiabetic;
CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
CC encoding them can be used in gene therapy, antisense therapy and vaccine
CC production. The proteins and polynucleotides are useful for screening for
CC agonists or antagonists of a protein and for the treatment and diagnosis
CC of disorders associated with the activity of a protein e.g. inflammation,
CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
CC anaemia, platelet disorders, thrombocytopenia, wounds, burns, ulcers,
CC osteoporosis, severe combined immunodeficiency, eczema, allergic
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC neurological disorders.
XX
SO Sequence 2950 BP; 915 A; 592 C; 628 G; 815 T; 0 other;

Query Match	62.5%;	Score 1920.6;	DB 22;	Length 2950;
Best Local Similarity	99.8%;	Pred. No. 0;		
Matches 1923; Conservative	0;	Mismatches 4;	Indels 0;	Gaps 0;

QY	1146	tgaaagfcaactcaatctggagagcccttgagacaactctgtggagagcccaagctctggagagtgct	1205
Db	1	tgcaagfgaactcaatctggagagccctggagacaactctgtggataactcaatctctgtggagtgctg	60
QY	1206	ctccccaagatttgattctcgaaactctgagagatgtgtctcagaagctgaaatggagatgtctctc	1265
Db	61	ctcccaagaatttgattctcgaaactctgagagatgtgtctcagaagctgaaatggagatgtctctc	120
QY	1266	gacaactcgggctcctctgttaataatacaagctcaagaagttcaagccaagtataatctt	1325
Db	121	gacaactcgggctcctctgttaataatacaagctcaagaagttcaagccaagtataatctt	180
QY	1326	tcacaagttcatctccagagagatcacacagacaagacaagactcaagcttatttgaagcttca	1385
Db	181	tcacaagttcatctccagagagatcacacagacaagacaagactcaagcttatttgaagcttca	240
QY	1386	tgagccagagagaggttgagccaagagggaaatgttactctgcagaanaaaygtttccattcga	1445
Db	241	tgagccagagagaggttgagccaagagggaaatgttactctgcagaanaaaygtttccattcga	300
QY	1446	catatacatcaactatataagcagctgtctccggtacacccgtgtgtcaatctgtggagaagcac	1505
Db	301	catatacatcaactatataagcagctgtctccggtacacccgtgtgtcaatctgtggagaagcac	360
QY	1506	caaggcctgttatgtgaagacacctgcgaagagtgatacaacacggctgcctctctcgagcttc	1565
Db	361	caaggcctgttatgtgaagacacctgcgaagagtgatacaacacggctgcctctctcgagcttc	420
QY	1566	catctgcagagagcctctctctgtgagacaggaatcttgcagaagtgtgaaanaacacactga	1625
Db	421	catctgcagagagcctctctctgtgagacaggaatcttgcagaagtgtgaaanaacacactga	480
QY	1626	gcaagaagaattcttgaagccaataacataactcccttggagagtgatggatcacacttata	1685
Db	481	gcaagaagaattcttgaagccaataacataactcccttggagagtgatggatcacacttata	540
QY	1686	tcaagagagatgatacccaaatccagacccttgagccaagaaatttgaagcttctcttcaagtaa	1745
Db	541	tcaagagagatgatacccaaatccagacccttgagccaagaaatttgaagcttctcttcaagtaa	600
QY	1746	aagcttatataatacaatccagaggaatacccgattactatttgaactctcttgaacatt	1805
Db	601	aagcttatataatacaatccagaggaatacccgattactatttgaactctcttgaacatt	660
QY	1806	gccaactctgtgcagaatgtctctgtgaactcaataactcgaagcttcttggggagactatgc	1865
Db	661	gccaactctgtgcagaatgtctctgtgaactcaataactcgaagcttcttggggagactatgc	720
	1866	tctatgggaaaaagagctctgcagaagaacacaggttggaatcccaatggaagaagggcccagaac	1925
Db	721	tctatgggaaaaagagctctgcagaagaacacaggttggaatcccaatggaagaagggcccagaac	780
QY	1926	ctaaattcccaagcagaagctgtatacttctgttcttcaactggaagcagaatctcagaactct	1985
Db	781	ctaaattcccaagcagaagctgtatacttctgttcttcaactggaagcagaatctcagaactct	840
QY	1986	ggaagtcacactccggagatttcagacaagtctgataagaacagatacaatactcgggaa	2045
Db	841	ggaagtcacactccggagatttcagacaagtctgataagaacagatacaatactcgggaa	900
QY	2046	aataatcagctctgcacaagaagcctcagagctctgtaanaataaagagatgtgtgtgtgcctg	2105
Db	901	aataatcagctctgcacaagaagcctcagagctctgtaanaataaagagatgtgtgtgtgcctg	960
QY	2106	aagcctcaagtttgctctcagaacacctgataagaacattatctctcagatgctgtgagaagcac	2165
Db	961	aagcctcaagtttgctctcagaacacctgataagaacattatctctcagatgctgtgagaagcac	1020

QY	2166	tcocctcaaccaataagaatgaagagcgacatcaaatctgtfatacaaaacctgnaaaaccttgag	2225
DB	1021	tcocctcaaccaataagaatgaagagcgacatcaaatctgtfatacaaaacctgnaaaaccttgag	1080
QY	2226	tattatcaaccttaagaatcaaacgctgcgcgggtggctcgaactgaaacttgagaaact	2285
DB	1081	kattcaatgaaccttaagaatcaaacgctgcgcgggtggctcgaactgaaacttgagaaact	1140
QY	2286	gaagaaccttaacaaagctcctaataatgatacaataagaatgaaatgaagaagaatgctataaa	2345
DB	1141	gaagaaccttaacaaagctcctaataatgatacaataagaatgaaatgaagaagaatgctataaa	1200
QY	2346	actgagctgaagcgcttgnaaaaacctgnaagaagaatggtttatttaatttgaccocactgttc	2405
DB	1201	actgagctgaagcgcttgnaaaaacctgnaagaagaatggtttatttaatttgaccocactgttc	1260
QY	2406	tgacattgnaagaaggaatgtgattacatagatcaagctctcgtcaatggaacctgagact	2465
DB	1261	tgacattgnaagaaggaatgtgattacatagatcaagctctcgtcaatggaacctgagact	1320
QY	2466	tgaagaaatltcaatltagctccctgcctgtcctgtctgtctgcnaaatgcaatgaaactcagctca	2525
DB	1321	tgaagaaatltcaatltagctccctgcctgtcctgtcctgtcctgcnaaatgcaatgaaactcagctca	1380
QY	2526	gaactctcaacattgtgtcaaaactgagcaattcttgattatcaagaanaattacctgnaaaa	2585
DB	1381	gaactctcaacattgtgtcaaaactgagcaattcttgattatcaagaanaattacctgnaaaa	1440
QY	2586	agatggaatataagctctctcatagactgatacgacaagaatgaaacgctgcagagaacagctaac	2645
DB	1441	agatggaatataagctctctcatagactgatacgacaagaatgaaacgctgcagagaacagctaac	1500
QY	2646	cgacactgaatgctgcgcctgcgggctgtgacgtgtgcnaagcgagcctgagcaagcctgtgtgnaaaa	2705
DB	1501	cgacactgaatgctgcgcctgcgggctgtgacgtgtgtgcnaagcgagcctgagcaagcctgtgtgnaaaa	1560
QY	2706	tttggagagaagctcccaacaactgttcaagctcttggtttgnaaaaactggaacaactcaagaataac	2765
DB	1561	tttggagagaagctcccaacaactgttcaagctcttggtttgnaaaaactggaacaactcaagaataac	1620
QY	2766	agagattagaatttttaagtgacatttttctgnaagaagaaacctcgtgnaaaaactccagagatt	2825
DB	1621	agagattagaatttttaagtgacatttttctgnaagaagaaacctcgtgnaaaaactccagagatt	1680
QY	2826	gaatttgcgcggnaaatcgtgtgtgacgagtgatgatatgctgtgccttcatactgaggtgtatttga	2885
DB	1681	gaatttgcgcggnaaatcgtgtgtgacgagtgatgatatgctgtgccttcatactgaggtgtatttga	1740
QY	2886	gaactctaagaagaatttagtgtttttttgacttgacttaagaagaatttactactgatacagc	2945
DB	1741	gaactctaagaagaatttagtgtttttttgacttgacttaagaagaatttactactgatacagc	1800
QY	2946	attgatacagaanaacttgcgcaagtgattatccaagttaaaactttcttcgcaagaagcttaggct	3005
DB	1801	attgatacagaanaacttgcgcaagtgattatccaagttaaaactttcttcgcaagaagcttaggct	1860
QY	3006	tgcttggtgtgcgaatttgatgtatgtatgatactccagtggtattatcaagtgctctttaaactagt	3065
DB	1861	tgcttggtgtgcgaatttgatgtatgtatgatactccagtggtattatcaagtgctctttaaactagt	1920
QY	3066	aactgctc 3072	
DB	1921	aactgctc 1927	

```

RESULT      5
AAH34171/C
ID          AAH34171 standard; cDNA; 2735 BP.
XX
XX          AAH34171;
AC
XX
XX          03-SEP-2001 (first entry)
DT
XX

```


CC Include: breast cancer, disorders of development, inflammatory diseases
CC of the breast, fibrocystic changes, proliferative breast disease and
CC non-carcinoma tumours.

CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 421 BP; 126 A; 69 C; 77 G; 149 T; 0 other:

Query Match

7.9%; Score 242; DB 22; Length 421;

Best Local Similarity 100.0%; Pred. No. 3.8e-61; Mismatches 0; Indels 0; Gaps 0;

Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2781 aggtgcatcttttggaagaaccctcgaagaactccagcaggttggaattggcggaata 2840

180 aggtgcatcttttggaagaaccctcgaagaactccagcaggttggaattggcggaata 239

QY 2841 tcgtgtgagcagtgatgagcttgccctcatgggtgattggaatcctaagaact 2900

Db 240 tcgtgtgagcagtgatgagcttgccctcatgggtgattggaatcctaagaact 299

QY 2901 agtgtttttgacctagactaagaattctaccctgatacagcattagtcagaact 2960

Db 300 agtgtttttgacctagactaagaattctaccctgatacagcattagtcagaact 359

QY 2961 tagccaagtgatatacctaattctctgcaagaagctagctgttggtggaact 3020

Db 360 tagccaagtgatatacctaattctctgcaagaagctagctgttggtggaact 419

QY 3021 tg 3022

Db 420 tg 421

RESULT 9

AA123590

ID AA123590 standard; DNA; 220 BP.

AC AA123590;

DT 12-OCT-2001 (first entry)

DE Probe #13523 for gene expression analysis in human cervical cell sample.

XX Probe; human; microarray; gene expression; cervical epithelial cell;

XX cervical cancer; ss.

OS Homo sapiens.

XX WO200157278-A2.

PN 09-AUG-2001.

PD 30-JAN-2001; 2001WO-US00670.

PF 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488901/53.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human cervical epithelial cells -

PS Claim 25; SEQ ID No 13523; 487bp; English.

XX

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

Sequence 220 BP; 58 A; 32 C; 54 G; 76 T; 0 other;

Query Match

7.2%; Score 220; DB 22; Length 220;

Best Local Similarity 100.0%; Pred. No. 8.1e-55; Mismatches 0; Indels 0; Gaps 0;

Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2830 ttggcggggaatcgtgtgagcagtgatgagcttgccctcatgggtgattggaat 2889

Db 1 ttggcggggaatcgtgtgagcagtgatgagcttgccctcatgggtgattggaat 60

QY 2890 cttagaacttagtgttttttgaacttagtactaagaattctaccctgatacagaata 2949

Db 61 cttagaacttagtgttttttgaacttagtactaagaattctaccctgatacagaata 120

QY 2950 gtccagaacttagcccaagtgtatcccaagttacttctgcaagaagctagctgtt 3009

Db 121 gtccagaacttagcccaagtgtatcccaagttacttctgcaagaagctagctgtt 180

QY 3010 ggggtgcaatttgatgatgatgatcctcagtgattatacag 3049

Db 181 ggggtgcaatttgatgatgatgatcctcagtgattatacag 220

RESULT 10

AA148904

ID AA148904 standard; DNA; 220 BP.

AC AA148904;

DT 17-OCT-2001 (first entry)

DE Probe #17590 used to measure gene expression in human placenta sample.

XX Probe; microarray; human; placenta; antenatal diagnosis;

XX genetic disorder; ss.

OS Homo sapiens.

XX WO200157272-A2.

PN 09-AUG-2001.

PD 30-JAN-2001; 2001WO-US00663.

PF 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-48897/53.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human placenta cells -

PT analyzing gene expression in human placenta -
XX
XX
PS Claim 25; SEQ ID NO 17590; 654bp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 220 BP; 58 A; 32 C; 54 G; 76 T; 0 other;

Query Match 7.28; Score 220; DB 22; Length 220;
Best Local Similarity 100.0%; Pred. No. 8.1e-55;
Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2830 ttggcgggaaatcgctgagcagatgagtggttccttcacaggggtgatttgaagaat 2889
Db 1 ttggcgggaaatcgctgagcagatgagtggttccttcacaggggtgatttgaagaat 60

QY 2890 cctaagcaattagtggttttttgacttaagtaagaatttctactgacacagcatla 2949
61 cctaagcaattagtggttttttgacttaagtaagaatttctactgacacagcatla 120

QY 2950 gtcaagaaacttagccaagtgttataccaagttaactttctgcagaagaagctagctgtt 3009
121 gtcaagaaacttagccaagtgttataccaagttaactttctgcagaagaagctagctgtt 180

QY 3010 ggggtggcaattgattgattgattgattcctcaagtgatttatacag 3049
Db 181 ggggtggcaattgattgattgattgattcctcaagtgatttatacag 220

RESULT 11
AA109206
ID AA109206 standard; DNA: 220 BP.
XX
AC AA109206;
XX
DT 09-OCT-2001 (first entry)
XX
DE Probe #9197 used to measure gene expression in human breast sample.
XX
KW Probe; human; breast; disease; breast cancer; development disorder; ss;
XX inflammatory disease; proliferative breast disease; non-carcinoma tumour.
OS Homo sapiens.
XX
PN WO200157270-A2.
XX
XX 09-AUG-2001.

XX 29-JAN-2001; 2001WO-US00661.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-476286/51.
XX
XX Novel single exon nucleic acid probe used to measuring gene expression
XX in a human breast -
PS Claim 25; SEQ ID NO 9197; 322bp; English.

XX
CC The present invention relates to novel single exon nucleic acid probes.
CC The present sequence is one such probe. The probes are useful for
CC measuring human gene expression in a human breast sample, where the probe
CC hybridises at high stringency to a nucleic acid expressed in the human
CC breast. The probes are useful for predicting, diagnosing, grading,
CC staging, monitoring and prognosing diseases of the human breast,
CC particularly those diseases with polygenic aetiology. The diseases
CC include: breast cancer, disorders of development, inflammatory diseases
CC of the breast, fibrocystic changes, proliferative breast disease and
CC non-carcinoma tumours.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).
XX
SQ Sequence 220 BP; 58 A; 32 C; 54 G; 76 T; 0 other;

Query Match 7.28; Score 220; DB 22; Length 220;
Best Local Similarity 100.0%; Pred. No. 8.1e-55;
Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2830 ttggcgggaaatcgctgagcagatgagtggttccttcacaggggtgatttgaagaat 2889
Db 1 ttggcgggaaatcgctgagcagatgagtggttccttcacaggggtgatttgaagaat 60

QY 2890 cctaagcaattagtggttttttgacttaagtaagaatttctactgacacagcatla 2949
61 cctaagcaattagtggttttttgacttaagtaagaatttctactgacacagcatla 120

QY 2950 gtcaagaaacttagccaagtgttataccaagttaactttctgcagaagaagctagctgtt 3009
121 gtcaagaaacttagccaagtgttataccaagttaactttctgcagaagaagctagctgtt 180

QY 3010 ggggtggcaattgattgattgattgattcctcaagtgatttatacag 3049
Db 181 ggggtggcaattgattgattgattgattcctcaagtgatttatacag 220

RESULT 12
AAF58252
ID AAF58252 standard; DNA: 936 BP.
XX
AC AAF58252;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1835.
XX
XX Electron-transfer group; EFM; mismatch; genotyping;
XX gene expression; ss.
OS Synthetic.
XX
PN WO200107665-A2.
XX
XX 01-FEB-2001.
XX
XX 26-JUL-2000; 2000WO-US20476.
XX
XX 26-JUL-1999; 99US-0145695.
XX 17-MAR-2000; 2000US-0190259.
XX
XX
XX (CLIN-) CLINICAL MICRO SENSORS INC.
XX
XX Umek RM;
XX
XX WPI; 2001-159728/16.
XX
XX Nucleic acids containing electron-transfer group, useful as labels in
XX hybridization assays, e.g. for genotyping, allowing repeat analyses on
XX a single surface -
XX

[illegible][illegible]

Query Match	Score	DB	Length
3.0%;	Score 91.8;	DB 22;	Length 936;

[illegible]

PD 01-FEB-2001.
XX 26-JUL-2000; 2000MO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI: 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
XX Example 6; Page 128; 159pp; English.

XX The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
XX

SQ Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;

Query Match 3.0%; Score 91.8; DB 22; Length 936;
Best Local Similarity 0.8%; Pred. No. 1.6e-16;
Matches 6; Conservative 452; Mismatches 309; Indels 0; Gaps 0;

OY 2280 taacttgaagaacctacaaagctcataatgataacataaagatgaatgaagaatgc 2339
DB 17 www. 76
OY 2340 tataaactagctgaagccctgaaaaacctgaagaagatggttatttcattgaccca 2399
DB 77 www. 136
OY 2400 ctgtctgacattgagaggggaatgattacataagtcctctgcaagtgaacctg 2459
DB 137 www. 196
2460 tgaacttgaagaaltcaatagctcctgctgctgctgcaaatgcagtgaaatcct 2519
DB 197 www. 256
OY 2520 agctcagaatctcccaattggtcaaacctgacatctctgattatcagaatattcct 2579
DB 257 www. 316
OY 2580 ggaaaaagatggaatgaagctctcattgaactgatacgacagatgaacgtgtagaaca 2639
DB 317 www. 376
OY 2640 gctcaccgaactgattgctgcctgggctgtgacgtgcaaggcagcctgagcagcctgtt 2699
DB 377 www. 436
OY 2700 gaacatttggaggggtcccccacaactcgtcaagcttggttgaaaaactgagactcac 2759
DB 437 www. 496
OY 2760 agatacagagattagatttagtcatlttttgaaagaacctctgaaaactcca 2819
DB 497 www. 556
OY 2820 gcagttgaatttgcgggaaatcgtgtgagcagtgatgattggtccttcattggtgt 2879
DB 557 www. 616

OY 2880 atttgaagaatctaagcaattagtggttttttgacttagtactaaagaatttctaccga 2939
DB 617 www. 676
OY 2940 tccagcattagtcagaaanaacttagccaagtgltalccaaagttaactttctgcaagaagc 2999
DB 677 www. 736
OY 3000 taggcttggtgggtgcaatttcatgatatctcagtgltatca 3046
DB 737 www. 783

Search completed: March 25, 2002, 11:45:30
Job time: 3314 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2002, 11:37:02 ; Search time 2764.94 Seconds

(without alignments)
11939.149 Million cell updates/sec

Title: US-09-697-089-3

Sequence: 3072
1 atgaattcataagagacaa.....ctttaactagtaactgct 3072

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estlin:*
4: em_estom:*
5: em_estopl:*
6: em_estda:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: gb_estcl:*
11: gb_estc2:*
12: gb_hic:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rtd:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
C 1	551	17.9	552	13	AO309404 CITBI-EI-
C 2	466.4	15.2	480	10	AV719179
C 3	407.8	13.3	602	13	AZ360053
C 4	347	11.3	371	10	AV656315
C 5	285.8	9.3	476	11	AI263294
C 6	282.8	9.2	470	13	AO624020
C 7	223.8	7.3	404	13	AO889169
C 8	223	7.3	261	10	AW337918
C 9	170	5.5	553	13	AO320928
C 10	157.2	5.1	219	13	AO283886
C 11	153.8	5.0	251	11	BF903662
C 12					IL2-MT018

C 13	146	4.8	630	13	AO112439	AO112439 CIT-HSP-2
C 14	139.8	4.6	840	11	BF207840	BF207840 601862546
C 15	136	4.4	509	10	AW418826	AW418826 ha21e11.x
C 16	106	3.5	728	13	AZ720059	AZ720059 RPT-24-8
C 17	98.8	3.2	412	11	H25984	H25984 Y156g07.r11
C 18	93.4	3.0	499	10	AI023795	AI023795 0x08d03.x
C 19	66.2	2.2	429	13	AZ484615	AZ484615 IM0311P08
C 20	55	1.8	635	13	B58691	B58691 CIT-HSP-201
C 21	49.2	1.6	668	13	AZ762115	AZ762115 IM0556N15
C 22	46.8	1.5	625	13	AZ614134	AZ614134 IM0442N17
C 23	44.8	1.5	546	13	AZ362463	AZ362463 IM0107N03
C 24	42.2	1.4	1023	10	BE306291	BE306291 601103781
C 25	40.8	1.3	835	10	BE643259	BE643259 Cr12_8.FO
C 26	40.6	1.3	361	10	AW504276	AW504276 UT-HF-BNO
C 27	40.2	1.3	783	11	BC563851	BC563851 602584655
C 28	40	1.3	424	13	AO638259	AO638259 927P1-17A
C 29	40	1.3	518	10	AA669206	AA669206 ad93b07.s
C 30	39.8	1.3	939	11	BE887277	BE887277 601508592
C 31	39.6	1.3	467	10	AW988289	AW988289 u05f04.y
C 32	39.4	1.3	363	10	BE544063	BE544063 601069826
C 33	39.4	1.3	523	11	BF259332	BF259332 HVSMEF001
C 34	39.4	1.3	549	11	BG109381	BG109381 602280435
C 35	39.4	1.3	573	11	BF257885	BF257885 HVSMEF001
C 36	39.4	1.3	611	11	BF253881	BF253881 HVSMEF000
C 37	39.4	1.3	754	11	BF258252	BF258252 HVSMEF001
C 38	39.4	1.3	854	11	BE881131	BE881131 601492001
C 39	39.2	1.3	759	10	AI729528	AI729528 BNGH1135
C 40	39.2	1.3	806	13	AO481970	AO481970 GNE1662.m
C 41	39.2	1.3	1101	13	CNS016LW	AL106910 Drosoph11
C 42	39	1.3	1201	13	CNS007JU	AL067358 Drosoph11
C 43	38.6	1.3	574	10	AW658668	AW658668 95275.MAR
C 44	38.4	1.2	425	10	AI358206	AI358206 gV96907.x
C 45	38.4	1.2	640	11	BI391271	BI391271 ppp1n.pk0

ALIGNMENTS

RESULT 1
LOCUS AO309404 552 bp DNA
DEFINITION CITBI-EI-2528J13.TF CITBI-EI Homo sapiens genomic clone 2528J13,
DNA sequence.
ACCESSION AO309404.1 GI:4041438
VERSION
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 552)
Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
Venter,J.C.
TITLE
Use of a random human BAC End Sequence Database for Sequence-Ready
Map Building
JOURNAL
Unpublished (1998)
COMMENT
Other-GSSs: CITBI-EI-2528J13.TF
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are available from Research Genetics (Inforesgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.
Location/Qualifiers
1..552
/organism="Homo sapiens"

FEATURES
source

```

/db_xref="taxon:9606"
/clone_1="2528J13"
/clone_1b="CITBI-E1"
/sex="male"
/cell_type="sperm"
/note="Vector: pBeloBAC11; Site_1: EcoRI; Site_2: EcoRI;
Caltech Human BAC Library D"
BASE COUNT      144 a      121 c      130 g      157 t
ORIGIN

Query Match      17.9%; Score 551; DB 13; Length 552;
Best Local Similarity 100.0%; Pred. No. 1.2e-135;
Matches 551; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1427 aaatggttccatttcgacattacacactatagcagctgtcgcgtacacctgtg 1486
Db 552 AAATGTTTCATTTCGGGACATTACATCCATTATAGCAGCTGCTCGGTACACCTGTG 493

Oy 1487 ggtcactgtggaagccaccagggctgtatgaagacctcgacagcagtgtalcaacag 1546
Db 492 GGTCAATCTGTGGAGACCAACAGGGGCTTATGAGCACCTCCGACAGGTATCAACACG 433

Oy 1547 gctgctctcgtgacttccatcgccaaggcctctctgtgagacaggaaatttgcaca 1606
Db 432 GCTGCTCTCTCGACTCTTCATCGCCAGAGGCTCTCTGTGGAGACAGGAACTTTGCCAA 373

Oy 1607 gtgtgaaaaaacacacacgtgacagaagaattctgaagccataacatcctctgtag 1666
Db 372 GGTGTAAGAAACACACACGTGAGCAAGAAATCTGAAGCCATAACATCTCTTTGTAG 313

Oy 1667 agtgtgcatccattatatacaagagatgacatccaatcagccctgagccaagaattg 1726
Db 312 AGTGTGGCACTCCATTATATCAAGAGAGTACATCCAAATCAGCCCTGAGCAAGAATTGG 253

Oy 1727 aagcttctctcaagggttaaaagcttatatacaactcagggaacatccccgatttactat 1786
Db 252 AAGCTTTCTTTCAAGGTAAACCTTATATATCAACTCAGGGAACATCCCGATTAATTAT 193

Oy 1787 ttgacttctcttgaacatttgcacaaattgtgcagtgctctgagctcatttaaaactgagct 1846
Db 192 TTGACTTCTTTGACATATTGGCCCAATTGTGCAAGTCTCTGAGACTCATTAACATGCACT 133

Oy 1847 ttatgtggggagactatgctcattatggaagaaggctgcagagaacacaggttgaaatccaca 1906
Db 132 TTTATGGGGGAGACTATGCTTCATGGGAAAGGCTGCAGAAACACAGGTGGAATCCACA 73

Oy 1907 tgggaagagagcccccagaacactattccagcagagggttatcttgtcttcaactgga 1966
Db 72 TGGAAAGAGGCCCCAGAAACCTACATTCCAGCAGAGGCTGTATCTTGTCTCAACTGGA 13

1967 agcaggaattc 1977
Db 12 AGCAGGAATTC 2

RESULT 2
AV719179/c 480 bp mRNA EST 16-OCT-2000
LOCUS AV719179 GLC Homo sapiens cDNA clone G1CE0A10 5', mRNA sequence.
DEFINITION AV719179
ACCESSION AV719179
VERSION AV719179.1 GI:10816331
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 480)
AUTHORS Qian,B., Wu,T., Huang,Q., Kang,B., Gao,X., Xu,Z., Xiao,H.,
Xu,X., Li,N., Peng,Y., Liu,F., Ou,J., Song,H., Cheng,Z., Zeng,L.,
Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Yang
Y., Gu,Y., Chen,Z. and Han,Z.
TITLE Homo sapiens cDNA GLC clones

```

```

JOURNAL Unpublished (2000)
COMMENT Contact: Zenguan Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES
Source 1..480
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1="GLCE0A10"
/clone_1b="GLC"
/tissue_type="Corresponding non cancerous liver tissue"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT      138 a      120 c      85 g      137 t
ORIGIN

Query Match      15.2%; Score 466.4; DB 10; Length 480;
Best Local Similarity 99.8%; Pred. No. 4e-113;
Matches 467; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 2605 catgaactatcagacaggttgacgctgtagaacacgtcacccgactgtagtgcacctg 2664
Db 480 CATGAACCTATCAGACAGATGAAAGTGTCTAGAACACCTCACCCACTGATGCTGCCCTGG 421

Oy 2665 ggcctgtgagctgcaaggcagcctgtagcagcctgttgaacatttggagaggtccacaa 2724
Db 420 GGCTGTACCTGCAAGAGCAGCGCTGAGCAGCCTGTGTAACATTTGGAGAGGTGCCACAA 361

Oy 2725 ctgcgtcaagcttggtgttgaaaaacttgagacacacagatacagagattgaatttagt 2784
Db 360 CTGCTCAACCTTGGGTTGAAAACTGTGAGACCTCACAGATACAGAGATTGAAATTTAGGT 301

Oy 2785 gcaatttcttgaagaacacctctgaaaaacttccagcagttgaatttgcgaggaactgt 2844
Db 300 GCATTTTGGAAAGAACCTCTGANAACCTTCACACAGTTGAATTTGGCGGAATTCGT 241

Oy 2845 gtgagcagtgatgagatggttgccttcacatggtgttatttgaagaatcctaagaacttagt 2904
Db 240 GTGAGCAGTGATGAGATGGCTTGCTTCATGGGTGTATTGAGAACTTTAAGCAATTAAGTG 181

Oy 2905 tttttgactttagtactaagaatttcaactgataccagcatagtcagaaacttagc 2964
Db 180 TTTTGTGACTTTAGTACTAAGAAATTTCTACCTGATCCAGCATTAAGTACGAAAACTTAGC 121

Oy 2965 caagtgtaaccagtttaacttctgcagaagaagcctgttgggtggcaatttagt 3024
Db 120 CAGGTGTATTCAGATTACTTTCTGCAGAGAGCTAGAGCTGTGTGGTGCGCAATTGAT 61

Oy 3025 gatgatgacatcagtgtaattacaagtgctttaaactgtaactgct 3072
Db 60 GATGATGATCTCAGTGTATTATACAGGAGCTTTAAACTGATTAACCTGCT 13

RESULT 3
AZ360053/c 602 bp DNA GSS 02-OCT-2000
LOCUS AZ360053
DEFINITION 1M0103H1F Mouse 10kb plasmid U08C1M library Mus musculus genomic
clone U08C1M0103H1 F, DNA sequence.
ACCESSION AZ360053
VERSION AZ360053.1 GI:10473753
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.

```


Qy	1993	ccccgcagggcgtctgtatcttcttgcttcctcaactcgtgaagcaggaattcgaagcctctgtgagtc	1992
Db	61	CCcAGcAGGGGcGTGTATCTTTGTCTTCACCTGGAAcCAGAAATTCAGGACTCTGGAGGTC	120
Qy	1993	acaatccggagatttcagcaagcttgaataaacaagatatacaatattctggagaaatattc	2052
Db	121	ACACTCCGGGGATTTCACCAAGTGTGAATTAACCAAGATATCAGATATCTGGGAAAAATATTC	180
Qy	2053	agctctgcacaaagcctccagcctcgcacataaagaatctgctcggctgtgcctgtgaagcctc	2112
Db	181	AGCCTCTCCCAAAACCCCTCAGGCTGCGCAATTAAGATGTGCTGTGGCTGGAAACCTTC	240
Qy	2113	agtttctgcctcagcacactcgttaagaacattctcttcctcaatgctgaagcagcctccctc	2172
Db	241	AGTTTGGTCCCTCAGCACACTGTGAAGAACATTATTTCTCTCATGTGTGAACCAACGATCCCTTC	300
Qy	2173	accatagaagatctgagagcacaatcacaatctctgtataacaacctgtaaaccttga	2227
Db	301	ACCATGAAGAAGTGTGAGGCGACATCTACATCTGTGTAAACAACCTGGACAGCTTGGGGTA	355

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
AI1263294/c				
LOCUS				
DEFINITION	AI1263294	364 bp	mRNA	EST
ACCESSION	qx37b01.x1			
VERSION	AI1263294			
KEYWORDS	AI1263294.1	GI:3871497		
SOURCE	EST.			
ORGANISM	human.			
	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Carnivora; Homnidae; Homo.			
	1 (bases 1 to 364)			
	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .			
	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),			
	Tumor Gene Index			
	Unpublished (1997)			
	Contact: Robert Strausberg, Ph.D.			
	Email: cgapbs-remail.nih.gov			
	Life Technologies catalog #: 11548-013			
	DNA Sequencing by: Washington University Genome Sequencing Center			
	Clone distribution: NCI-CGAP clone distribution information can be			
	found through the I.M.A.G.E. Consortium/LMW at:			
	www-bio.11nl.gov/bdnp/image/image.html			
	Insert Length: 2146	Std Error: 0.00		
	Seq primer: -400P from Glibco			
	High quality sequence stop: 364.			

ATTURES	SOURCE	Location/Qualifiers
		1. .364
		/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/clone="IMAGE:2005417"
		/clone_idb="NCI_CGAP_Pan1"
		/tissue_type="adenocarcinoma"
		/lab_host="DH10B"
		/note="Organ: pancreas; Vector: PCMV-SPOrt6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.72 kb. Life Technologies catalog #: 11548-013"
BASE COUNT	117 a	55 g 108 t
ORIGIN	84 c	

Query Match	11.1%	Score 342	DB 10	Length 364
Best Local Similarity	100.0%	Pred. No. 5.1e-80		
Matches 342	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY_2731	aagcttgagttggaanaacatgagactccagatcaagaagatattgagtcattc	2790		
364	AAGCTTGAGTTGGAANAACATGAGACTCCAGATCAAGAAGATATTGAGTCATT	305		

Qy	2791	tttgaagaagacccctcgaaaacctccagagctgtaatttcggcggaatcgtgtgcg	2850
Db	304	tttggaaagaacccctcgaaaacccctccagcagctgtaattttggcggaaatcgtgtgcg	245
Qy	2851	agtgatgatgcttcgtccctccacgggtgtaatttgagaactcttaagcaattagtctttt	2910
Db	244	agtgatgatgcttcgtccctccatggctgtaatttgagaactcttaagcaattagtctttt	185
Qy	2911	gacttgctgctctaaagaattcttacctgatacgaattgtagaanaactatgccaagt	2970
Db	184	gacttgctgctctaaagaattcttacctgatacgaattgtagaanaactatgccaagt	125
Qy	2971	ttaccacaagttaactttctccgcaagaagctagaagctttgtgggtggcaatttgatgat	3030
Db	124	ttatcccaagtttaacttttctgcagaagaagctagagctttgtgggtggcaatttgatgat	65
Qy	3031	gattcagtgcttattcaagtgctttttaactgtacagct	3072
Db	64	gattcagtgcttattcaagtgctttttaactgtacagct	23

RESULT	6
LOCUS	BG210375
DEFINITION	BG210375 476 bp mRNA EST 21-APR-2001
ACCESSION	R512913 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
VERSION	BG210375
KEYWORDS	BG210375.1 GI:13732062
SOURCE	EST.
ORGANISM	human.
REFERENCE	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 476)
	Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R., Cain,S., Dall,T., Thornton,M., Ramachandran,R., Whittington,J., Lerner,L., Krashoc,D., McElligott,K., Clark,S., Mays,R., Smith,E. Veloso,N., Hess,J., Cochren,K., Lo,K., Offenbacher,J., Danzig,J. and Ducar,M.
TITLE	Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression
JOURNAL	Nat. Biotechnol. 19 (5), 440 (2001) In press
COMMENT	Contact: Scott J. Cain Athersys, Inc. 3201 Carnegie Ave, Cleveland, OH 44115, USA Tel: 216 431 9900 Fax: 216 361 9596 Email: scai@atersys.com
FEATURES	High quality sequence stop: 360.
source	Location/Qualifiers 1..476

FEATURES	LOCATION/QUALIFIERS
SOURCE	1. .476
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone_lib="Athensys RAGE Library"
	/cell_line="HT1080"
	/note="See 'Creation Of Genome-Wide Protein Expression Libraries using Random Activation Of Gene Expression', Nature Biotechnology, in press. Note that even though 'cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."
BASE COUNT	120 a 99 c 107 g 149 t 1 others
ORIGIN	

	Query Match	Similarity	9.3%	Score 285.8	DB 11	Length 476
	Best Local	Similarity	97.6%	Pred. No. 4.9e-65		
	Matches	290	Conservative	0	Mismatches	7
					Indels	0
					Gaps	0
OY	2776	atttaagtgcgaatttttggagaagccctctgaaaaacttccagcagtttgtaatttggcg				2835
Db	145	ACTATAAGTGCATTTCTTGGAAGAAGAACCCCTCGAAAAAATCTTCACGACGTGTAATTTGGGCG				204
OY	2836	ggaatctgttgagcagtgatgcagcttcgttccttcatttggtgctatttgagaattcttaag				2895

```

|||||
Db 205 GCAATTCGTGACGATGATGATGCGCTTGCTGCTGCTGATTTGACATCTTAAG 264
Oy 2896 caataagtgcttttgacttgacttgacttgacttgacttgacttgacttgacttgact 2955
Db 265 CAATTAGTGTGTTTGTGACTTGTAAAGAAATTTCTACCTGATCCACATTAATGACAA 324
Oy 2956 aaacttaacaaagtgttaccgaagtacattcttcgaagaagcgaagctgtgtgtgtg 3015
Db 325 AAACCTAGCAGATGATCTGATCCAGTTACTTTCTGCAAGAAAGCTAGGCTTTGGGTGG 384
Oy 3016 caatttgatgatgatgatgatgatgatgatgatgatgatgatgatgatgatgatgatgat 3072
Db 385 CAATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 441

```

RESULT 7

AO624020 470 bp DNA GSS 16-JUN-1999
 DEFINITION HS_5378_B2_C12.SP6.RPCI-11 Human Male BAC Library Homo sapiens
 ACCESSION AO624020 genomic clone Plate=954 Col=24 Row=F, DNA sequence.
 VERSION AO624020.1 GI:5086412
 KEYWORDS GSS.

SOURCE

human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 470)
 Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
 Hood,L.

REFERENCE
 AUTHORS
 TITLE Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 JOURNAL 99380589
 MEDLINE
 COMMENT Contact: Mahairas GC, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones are derived from the human BAC library RPCI-11. For BAC
 library availability, please contact Pieter de Jong
 (pieter@dejong.med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
 or from Research Genetics (info@resgen.com). BAC end web Server:
 http://www.htsc.washington.edu
 Plate: 954 row: F column: 24
 Seq primer: SP6
 Class: BAC ends
 High quality sequence stop: 470.

FEATURES

source
 1..470
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=954 Col=24 Row=F"
 /clone_lib="RPCI-11 Human Male BAC Library"
 /sex="male"
 /note="Vector: pBACE3.6; Site: 1: EcoRI; Site: 2: EcoRI;
 Male blood DNA was isolated from one randomly chosen donor
 and partially digested with a combination of EcoRI and
 EcoRII. Size selected DNA was cloned into the
 pBACE3.6 vector at EcoRI sites"
 pBACE3.6 vector at EcoRI sites"
 BASE COUNT 142 a 101 c 107 g 117 t 3 others
 ORIGIN

Query Match 9.2%; Score 282.8; DB 13; Length 470;
 Best Local Similarity 84.1%; Pred. No. 3.1e-64;
 Matches 317; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

```

Oy 1972 gaattcagactcttgaggatcactccggatcttcagcaagttgaataagcaagatalc 2031
Db 2 CTAATCATGACTCTGGAGGTCTACACTCCGGGATTTGACGAACATTAAGCAAGATATC 61
Oy 2032 acatactgagggaataatcactcagctcgcacaagcctcagctcagctcagctcagctc 2091
Db 62 AGATATCTGGGGGAAATATATCTAGCTCTGCCACACCTTCAGGCTCAATAAAGAGATCT 121
Oy 2092 gctgggtgagctggaagcctcagctctgtgtgtcctcagcactgtgaagaatattctctc 2151
Db 122 GCTGGTGTGGCTGGGAAGCCTCAGTTGTGCTCAGCAGCAGCTGTAAGACCATTAATTCCTC 181
Oy 2152 atggtggaagcagctccctccacatagaagaaggagagacatcacatctglaacaac 2211
Db 182 ATGTGGGAAGCCAGTCCNCTACCATAGATGAGAGGACACATCATCTGTATACANNC 241
Oy 2212 ctgaaaccttgagatcattcattcactcactcactcactcactcactcactcactcactcact 2271
Db 242 CTGAAGAACCTTGAGTATCATGACCTACAGATCAACGCGCTCCGGTATTTGTAATATTC 301
Oy 2272 agcttggtgacttgagaagcctcactcactcactcactcactcactcactcactcactcact 2331
Db 302 AGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCT 361
Oy 2332 gaagatgctataaact 2348
Db 362 GAGGAGACTTATATCTCT 378

```

RESULT 8
 AO889169/c
 LOCUS HS_2161_B1_A01_77C.CIT Approved Human Genomic Sperm Library D Homo
 DEFINITION sapiens genomic clone Plate=2161 Col=1 Row=B, DNA sequence.
 ACCESSION AO889169
 VERSION AO889169.1 GI:6345359
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 404)
 Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
 Hood,L.

REFERENCE
 AUTHORS
 TITLE Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 JOURNAL 99380589
 MEDLINE
 COMMENT Contact: Mahairas GC, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones may be purchased from Research Genetics (info@resgen.com).
 BAC end web Server: http://www.htsc.washington.edu
 Plate: 2161 row: B column: 1
 Seq primer: T7
 Class: BAC ends
 High quality sequence stop: 404.

FEATURES

source
 1..404
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=2161 Col=1 Row=B"
 /clone_lib="CIT Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: sperm; Vector: pBelBAC11; BAC Clones in
 E-Coli DH10B"
 BASE COUNT 119 a 93 c 70 g 122 t
 ORIGIN

Query Match 7.3%; Score 223.8; DB 13; Length 404;
 Best Local Similarity 91.5%; Pred. No. 1.5e-48;
 Matches 248; Conservative 0; Mismatches 22; Indels 1; Gaps 1;

QY 2 tgaattcataaaggacataagccgagcccttattcaagaatggaatgactgtttaa 61
 ||||||| ||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 DB 350 TGAATTCATAGAGACAGATAGCCGAG-CCTTATTCATAAAGATGGATGCTGTATAG 292
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 QY 62 agcaaatcacagatgaccttttgatggaatgttcgaatcgaggaagtaaacatca 121
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 DB 291 AGCTACTCACAGATGACCTTTTGTATGAATGTTATGAATCGAAGAAAGTAACATCA 232
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 QY 122 ttctgctgagaaggtgagcagagatgctgtagaaggaatcatcatgatttgaaga 181
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 DB 231 TTTTCTGGAGAAAGGTGAGACAGATGCTGTAGAGGATCATTCACATGATTTGAAG 172
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 QY 182 aggggtcagaagctctgaacctcttcttaaaccttaaggagtggaactatctat 241
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 DB 171 AGGTTTCAGAGTCTCTGTAACCTCTATGTTAAATCCCTTAAGAGACTGAACATCTCTAT 112
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 DB 242 ttcaagactggaatgagcaaatgtcttttca 272
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 DB 111 TTCAGGACTTGAAGGACAGATGATATCA 81

RESULT 9
 AM337918 261 bp mRNA EST 31-JAN-2000
 LOCUS AM337918/c
 DEFINITION he12h11.x1 NCI-CGAP_CML1 Homo sapiens cDNA IMAGE:2918853 3',
 mRNA sequence.
 ACCESSION AM337918 GI:6834544
 VERSION AM337918
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE

AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 COMMENT Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs@email.nih.gov
 Tissue Procurement: Elisabeth Paletta, Jonathan D. Licht, M.D.,
 Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life
 Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The
 I.M.A.G.E. Consortium DNA Sequencing by: Washington University
 Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www-bio.llnl.gov/bdnp/image/image.html
 Seq primer: -40UP from Gibco
 High quality sequence stop: 201.
 Location/Qualifiers
 1. 261

FEATURES

source
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2918853"
 /clone_lib="NCI-CGAP_CML1"
 /tissue_type="myeloid cells, 18 pooled CML cases, BCR/ABL
 rearrangement positive, includes both chronic phase and
 myeloid blast crisis"
 /lab_host="DH10B"
 /note="Organ: whole blood; Vector: pCMV-SPORT6; Site:1:
 SalI; Site:2: NotI; Cloned unidirectionally. Primer:
 oligo dT. Library constructed by Life Technologies."
 BASE COUNT 86 a 54 c 38 g 83 t
 ORIGIN

Query Match 7.3%; Score 223; DB 10; Length 261;
 Best Local Similarity 100.0%; Pred. No. 2.3e-48;
 Matches 223; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2850 cagtgatgagtgcttccttcattgggtgattttggaatcctaagaattaggtttt 2909
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 DB 261 CAGTGATGATGCTTCCTTCATAGGCGTGTATTTGAGAAATCTTAAGCAATTAAGGTTT 202
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 QY 2910 tgaattagtaactaaagaatttcaactgattccagatctagtcagaactagcaagt 2869
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 DB 201 TGAATTTAGTACTTAAGAAATTTCTACTGATCCAGCATTTGTGAGAAACTTAGCCAGT 142
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 QY 2970 gtatccaaagttaactttcttgcagaagcagtgctgttgglygcaattgatgta 3029
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 DB 141 GTATTCGAAGTTAACTTTCTGCAAGAGCTAGGCTGTGGGCGCAATTTGATGATGA 82
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 QY 3030 tgatctcagtgattatacagtgcttttaactagtaagct 3072
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 DB 81 TGATCTCAGTGTATTAACAGTGCTTTTAACATGATACGTCT 39

RESULT 10
 AQ320928 553 bp DNA GSS 06-MAY-1999
 LOCUS AQ320928/c
 DEFINITION RPC111-93C9.TV RPC1-11 Homo sapiens genomic clone RPC1-11-93C9, DNA
 sequence.
 ACCESSION AQ320928
 VERSION AQ320928
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (Bases 1 to 553)
 AUTHORS Adams M.D., Rounsley S.D., Zhao S., Bass S., Linher K., Golden K.,
 Berry K., Granger D., Suh E., Wible C., de Jong P. and Venter J.C.
 TITLE Use of Human BAC End Sequences for Sequence-Ready Map Building
 JOURNAL Other GSSs: RPC111-93C9.TV
 COMMENT Contact: Shaying Zhao, William Nierman, Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: hbeef@igf.org

clones are derived from the human BAC library RPC1-11. For BAC
 library availability, please contact Pieter de Jong
 (pieter@dejong.med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
 Research Genetics (info@resgen.com). BAC end search page:
 http://www.tigr.org/db/humgen/bac_end_search/bac_end_search.html
 Seg primer: T7
 Class: BAC ends.
 Location/Qualifiers
 1. 553

FEATURES
 source
 /organism="Homo sapiens"
 /db_xref="GDB:753384"
 /db_xref="taxon:9606"
 /clone="RPC1-11-93C9"
 /clone_lib="RPC1-11-93C9"
 /sex="Male"
 /cell_type="Lymphocytes"
 /note="Vector: PBACE3.6; Site:1: EcoRI; Site:2: EcoRI;
 RPC11 Human Male BAC Library"
 BASE COUNT 170 a 107 c 114 g 162 t
 ORIGIN

Query Match 5.5%; Score 170; DB 13; Length 553;
 Best Local Similarity 100.0%; Pred. No. 3.4e-34;
 Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

0y 2615 tcgacagatgaacgtgcttagaagacagctcaccgcatgctgcccctggggctgtagc 2674
|||||
Db 258 TCGACAGATGAACGTGCTAGAAAGCTCACCGCACTGATGCTGCCCTGGGGCTGTGAGC 199
0y 2675 tgcgaagcagcctgagcagcctgctgttgaacatttgagagaggtcccaactcgtccaagc 2734
|||||
Db 198 TGCAGGAGCGAGCTGAGCGAGCTGTGAAACATTTTGAGAGAGTCCACAAACGTCGCAAGC 139
0y 2735 ttgggttcaaaaactgagagacacagatataagattgaattagct 2784
|||||
Db 138 TTGGGTTAAAAAAGCTGAGAGACTCACAGATACAGAGATTGAATTTTAGCT 89

RESULT 11
AQ28386
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AQ28386 219 bp DNA GSS 27-APR-1999
RPII11-78E13.TV RPI1-11 Homo sapiens genomic clone RPI1-11-78E13,
DNA sequence.
AQ28386
AQ28386.1 GI:3910204
GSS.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
Use of human BAC End Sequences for Sequence-Ready Map Building
Unpublished (1998)
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0200
Email: mdadams@ligr.org
Clones are derived from the human BAC library RPI1-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genetics (info@resgen.com). BAC end search page:
http://www.tlgr.org/tldb/humgen/bac_end_search/bac_end_search.html
Seq primer: 77
Class: BAC ends.

FEATURES
Source

1. 219
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="GDB:7529676"
/db_xref="taxon:9606"
/clone_lib="RPI1-11-78E13"
/clone_lib="RPI1-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
RPI11 Human Male BAC library"
BASE COUNT 69 a 45 c 46 g 59 t
ORIGIN

Query Match 5.1%; Score 157.2; DB 13; Length 219;
Best Local Similarity 86.1%; Pred. No. 7.5e-31;
Matches 174; Conservative 0; Mismatches 28; Indels 0; Gaps 0;
0y 2099 tggctggaagcctcagttgtgctcctcagcactgtaagaacattatctcctcagtgtag 2158
|||||
Db 1 TGGCTGGAAGCCTTAGTGTGCTCCTCAGACCTGTAAGAACAATTATCTCTCATAGTGTG 60
0y 2159 aagcgaagcctcctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2218
|||||
Db 61 AAGCCAGTCCTCCTGACCAAGAGATGAGAGGACATCATCATGTAACAAACCTGAAAA 120
0y 2219 ccttgatattcatgactacagatacaagcgtgcgggtgtgtcgtgactgacagccttgg 2278

Db 121 CCTTGAGTATTCATGACCTACAGAAATCAAGCGGCTGCGGGTATTGTTAATATCAATGCTG 180
0y 2279 gtaacttgaagaaccttaacaa 2300
|||||
Db 181 TGTGCTTGTTCACCTTAAAAA 202

RESULT 12
BF903662
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT

BF903662 251 bp mRNA EST 18-JAN-2001
IL2-MT0180-181200-276-F03 MT0180 Homo sapiens cDNA, mRNA sequence.
BF903662
BF903662.1 GI:12295121
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?fl=IL2&t2=IL2-MT0180-
181200-276-F03&t3=2000-12-18&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 45
High quality sequence stop: 96.

FEATURES
Source

1. 251
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MT0180"
/dev_stage="Adult"
/note="Organ: marrow; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
BASE COUNT 72 a 62 c 56 g 61 t
ORIGIN

Query Match 5.0%; Score 153.8; DB 11; Length 251;
Best Local Similarity 85.5%; Pred. No. 6.2e-30;
Matches 218; Conservative 0; Mismatches 32; Indels 5; Gaps 4;
0y 1007 agaccctctcttgygcacacactgycacaatcagatgggt-gaaagtgaagtcacac 1065
|||||
Db 1 AAGCCCTCTCTTTAGGTCATCATTTTGTGCAATCTCTATGGACAAAGGAGTTCCAC 60
0y 1066 tctcacacacaacaagcgtctccatccatctatgatcgtctgtatcacagaaaaacaa 1125
|||||
Db 61 TCTCACACACAACAACCGCTGATCATCTCTATGATGTGAAGATACAGAAAAAANA 120

QY 1126 cacaacataaagtgctgcaagtgaacttcgtggagccttgagaccactgtggaac 1185
|||||
Db 121 CACAACATTAAGGCGCTGTG--ATGTGACTCATTCGGAGCCT-GACCACTGTGG-CAC 176
QY 1186 ctgagcttggaagggtgtgtctcccaagtttgatctgaacgcaggtgtgtccagc 1245
|||||
Db 177 CTGATCTGTGGTGGTGCTTGTCCACAACTTGAAATTCGAATTCGACGAGATGTCCACCGC 236
QY 1246 gtgaatgagagatgctc 1260
|||||
Db 237 GCGTACTAGAAATGTC 251

RESULT 13
AQ112439 630 bp DNA GSS 29-AUG-1998
LOCUS CIT-HSP-2372C1.TR CIT-HSP Homo sapiens genomic clone 2372C1, DNA
DEFINITION sequence.
ACCESSION AQ112439
VERSION AQ112439.1 GI:3484599
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 630)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
Venter,J.C.
TITLE Use of a random human BAC End Sequence Database for Sequence-Ready
Map Building
JOURNAL Unpublished (1998)
COMMENT Other_GSSs: CIT-HSP-2372C1.TF
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page: http://www.tigr.org/cdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends.
FEATURES
Location/Qualifiers
1..630
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2372C1"
/clone_lib="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBelobAC11. Site_1: HindIII. Site_2:
HindIII"
BASE COUNT 188 a 126 c 124 g 192 t
ORIGIN
Query Match 4.8%; Score 146; DB 13; Length 630;
Best Local Similarity 100.0%; Pred.No.8.5e-28;
Matches 146; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2349 agctgaagcctgaanaacctgaagaagatgtgtttatttcattgaccacattgttga 2408
|||||
Db 485 AGCTGAAGGCTCAAAAACCTGAAGAAGATGTGTTTATTTCATTGACCCACTTGTCTGA 544
QY 2409 cattggaagaagaatgatatactcaagctctgttcaagtgaaccctgtgaaccttga 2468
|||||
Db, 545 CATTGCAAGAGGAGATGATACATAGTCAAGTCTGTCTGAAGTGAACCCGTGTGACCTTGA 604
QY 2469 agaataatcaatagctctccctgctgt 2494

Db 605 AGAATTCATTAATGATCTCTGCTGCT 630
|||||
RESULT 14
BF207840/c 840 bp mRNA EST 06-NOV-2000
LOCUS BF207840
DEFINITION 601862546F1 NIH_MGC_53 Homo sapiens cDNA clone IMAGE:4082074.5',
mRNA sequence.
ACCESSION BF207840
VERSION BF207840.1 GI:11101426
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgabs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at: image.llnl.gov
Plate: LICW44 row: 1 column: 11
High quality sequence stop: 636.
FEATURES
Location/Qualifiers
1..840
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4082074"
/clone_lib="NIH_MGC_53"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: bladder; Vector: pDNR-LIB (Clontech);
Site_1: SfiI (ggcgctcgcc); Site_2: SfiI (ggcattatggcc
); Double-stranded cDNA was prepared from cell line RNA.
5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGCCATTGTGCC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCGCGAGCGGCCACATG-dT(30)BN-3'
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.55 kb (range 0.9-4.0 kb). 15/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)."
BASE COUNT 276 a 143 c 163 g 256 t
ORIGIN
Query Match 4.6%; Score 139.8; DB 11; Length 840;
Best Local Similarity 91.9%; Pred.No.4e-26;
Matches 203; Conservative 0; Mismatches 12; Indels 6; Gaps 5;

QY 2853 tgatgtagctgccttcacatggtgattggaacttaagaacttagtcttttga 2912
|||||
Db 679 TGATGTAATGCTTCCCTTCATGSGTGTA-TTGAAATCTTAACCAATTAAGTGT-TTGA 622
QY 2913 cttagtactaagaatcttcaactgatacgaacttagtcaagaacttagcc-aagtgt 2971
|||||
Db 621 CTTTAGTACTAAAGAAATTTCTACTGATCA-CATTAGTCTGAAGAACTTAGCCAAAGTGT 563
QY 2972 tatccaagttaacttctcgaagaagctagctgtgtgtggtgcaattgagatgatg 3031
|||||
Db 562 TATCCAGTTAACTTTTCTCAAGAAAGCTTAGGCT-CCTTGGGTGCTTGAATGATGATG 505
QY 3032 atccagtgcttatacagtgctttaaacttagtaactgtct 3072
|||||
Db 504 ATTCAGTGTATTATACAGGTGCTTTTAACTAGTAAGTACTGCT 464

RESULT 15
AM418826/c

LOCUS AM418826 509 bp mRNA EST 09-FEB-2000
DEFINITION ha21e11.x1 NCI_CGAP_Kid12 Homo sapiens cDNA clone IMAGE:2874380 3',
mRNA sequence.

ACCESSION AM418826
VERSION AM418826.1 GI:6946758

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 509)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL COMMENT
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bdip/image/image.html
Seq primer: -400p from Glibco
High quality sequence stop: 461.
Location/Qualifiers

FEATURES
source
1..509
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2874380"
/clone_id="NCI_CGAP_Kid12"
/tissue_type="2 pooled tumors (clear cell type)"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pT7/3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP_Kid5 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneids 1322912-1325831, 1471368-1472903 and
1492104-1493255). Subtraction by Bento Soares and M.
Fatima Bonaldo."

BASE COUNT 161 a 99 c 80 g 169 t
ORIGIN

Query Match 4.4%; Score 136; DB 10; Length 509;
Best Local Similarity 100.0%; Pred. No. 3.8e-25;

Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2937 tgaatccagcattagtcagaaacttagcccaagtgatcccaagtaactttctcagaaga 2996
|||||
DB 509 tgaatccagcattagtcagaaacttagcccaagtgatcccaagtaactttctcagaaga 450
OY 2997 agctagcttgggtgcaatttgatgatgatcatcgaatgtattacagtgcttt 3056
|||||
DB 449 AGCTAGGCTTGTTGGTGCGCAATTGATGATGATCATCTCAAGTGTATTACAGGTGCTTT 390
|||||
OY 3057 taaactagtaactgct 3072
|||||
DB 389 TAAACTAGTAAGTCTGCT 374

Search completed: March 25, 2002, 11:37:10
Job time: 2815 sec

THIS PAGE BLANK (USPTO)